Scientific Networks and IDEX funding

Alberto Corsini

UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE
DROIT ET SCIENCES POLITIQUES ÉCONOMIQUES ET DE GESTION

Réseaux Scientifiques et Financement IDEX

Alberto CORSINI

GREDEG

Présentée en vue de l'obtention du grade de docteur en sciences économiques d'Université Côte d'Azur

Dirigée par : Patrick MUSSO / Michele PEZZONI

Soutenue le : 09 / 09 / 2022

Devant le jury, composé de :

Francesco LISSONI, Full professor, University of Bordeaux

Patrick MUSSO, Full professor, Université Côte d'Azur

Lionel NESTA, Full professor, Université Côte d'Azur

Michele PEZZONI, Associate Professor, Université Côte d'Azur

Valentina TARTARI, Associate Professor, Copenhagen Business School

Reinhilde VEUGELERS, Full professor, Katholieke Universiteit Leuven

Réseaux Scientifiques et Financement IDEX

Jury :

Président du jury :
Lionel NESTA, Full Professor (HDR), Université Côte d'Azur, France

Rapporteurs :
Francesco LISSONI, Full Professor (HDR), University of Bordeaux, France
Reinhilde VEUGELERS, Full Professor (HDR), Katholieke Universiteit Leuven, Belgium

Examniateurs :
Lionel NESTA, Full Professor (HDR), Université Côte d'Azur, France
Valentina TARTARI, Associate Professor (HDR), Copenhagen Business School, Denmark

Directeur de thèse :
Patrick MUSSO, Full Professor (HDR), Université Côte d'Azur, France

Co-directeur de thèse :
Michele PEZZONI, Associate Professor (HDR), Université Côte d'Azur, France

Scientific Networks and IDEX Funding

ABSTRACT IN ENGLISH

The study of factors influencing scientific knowledge production and the design of financial incentives that may stimulate it have become increasingly relevant among scholars and policymakers (Stephan, 2012). This thesis focuses on the role played by three of the key actors in knowledge production: Ph.D. students, researchers, and universities. First, I investigate how the Ph.D. students' scientific production and network are associated with the characteristics of the training environment, including funding availability. Then, I quantify how a government funding program addressed to promote university excellence (IDEX) affects researchers' outcomes. Finally, I compare the effects of competitive grant funding versus block funding on the impact of the resulting researchers' articles. The empirical analyses of the whole dissertation are based on the French case.

In the first chapter, I ask: what makes a productive Ph.D. student? Specifically, I investigate how the social environment to which a Ph.D. student is exposed during her training relates to her scientific productivity. I focus on how supervisor and peers' characteristics are associated with the student's publication quantity, quality, and co-authorship network. Unique to my study, I cover the entire Ph.D. student population of a European country for all the STEM fields analyzing 77,143 students who graduated in France between 2000 and 2014. I find that having a productive, mid-career, low-experienced, female supervisor who benefits from a national grant is positively associated with the student's productivity. Furthermore, I find that having few productive freshman peers and at least one female peer is positively associated with the student's productivity.

In the second chapter, I estimate the impact of the initiative of excellence (IDEX) funding program on a broad set of French researchers' outcomes such as publication productivity, collaboration networks, research interdisciplinarity, patenting, mentoring of Ph.D. students, and fundraising. Relying on a panel of 32,947 researchers in STEM disciplines observed between 2006 and 2015, I investigate the effect of being affiliated with universities that applied for IDEX and universities that were awarded IDEX. Moreover, I investigate the indirect effect of IDEX on researchers in non-applicant universities who collaborate with researchers in awarded universities. Using a difference-in-differences approach, I find that both applying for IDEX and being awarded IDEX enlarge the researchers' collaboration networks. Being awarded IDEX is particularly beneficial for boosting collaborations with other French universities and international collaborations. I also find positive indirect effects of IDEX on the collaboration networks of researchers in non-applicant universities.

In the third chapter, I compare the effectiveness of two research funding models: block funding and competitive funding. EU governments are increasingly relying on competitive grants to allocate research funding, replacing the traditional block funding used to support research. The literature aiming at quantifying the impact of funding models has not yet answered the question: is grant-funded research more impactful than block-funded research? In the French context, I compare the impact of 6,441 scientific articles resulting from competitive grants with that of 6,441 similar articles resulting from block funding. I rely on publication acknowledgments to retrieve the funding information and on citation data to assess publications' impact. I apply a probabilistic matching procedure to compare similar articles. I find that publications receiving the support of competitive grants obtain significantly more citations than those supported by block funding in the long run, while the difference is not statistically significant in the short run.

My dissertation offers important insights to policymakers in designing effective training and financing policies for science.

Keywords: Ph.D. students, Researchers' outcomes, Productivity determinants, IDEX funding, Grant funding, Block funding.

RÉSUMÉ EN FRANÇAIS

L'étude des facteurs influençant la production scientifique et la conception d'incitations financières pour la stimuler sont devenues fondamentales pour les décideurs politiques (Stephan, 2012). Cette thèse se concentre sur le rôle joué par trois acteurs de la production scientifique : les doctorants, les chercheurs et les universités. D'abord, j'étudie comment la production scientifique des doctorants est associée aux caractéristiques de l'environnement de leur formation. Ensuite, je quantifie comment un programme de financement gouvernemental visant à promouvoir l'excellence universitaire (IDEX) affecte les résultats des chercheurs. Enfin, je compare les effets du financement compétitif par rapport au financement en bloc sur l'impact des articles des chercheurs. Les analyses empiriques de l'ensemble de la thèse portent sur le cas français.

Le premier chapitre répond à la question suivante : qu'est-ce qui rend un doctorant productif ? J'étudie comment l'environnement social auquel une doctorante est exposée pendant sa formation est lié à sa productivité scientifique. Je me concentre sur les caractéristiques du superviseur et des pairs qui influencent la quantité et qualité des publications ainsi que les collaborations scientifiques de l'étudiant. Unique à mon étude, je couvre l'ensemble de la population doctorale d'un pays européen pour tous les domaines STEM en analysant 77,143 étudiants diplômés en France entre 2000 et 2014. Je trouve qu'avoir une femme superviseure qui est productive, à mi-carrière, peu expérimentée, et qui bénéficie d'une subvention nationale est positivement associé à la productivité de l'étudiante, ainsi que d'avoir peu de pairs juniors et productifs et au moins un pair de sexe féminin.

Dans le second chapitre, je quantifie l'impact du IDEX sur un large éventail de résultats de chercheurs français tels que la productivité et les collaborations scientifiques, l'interdisciplinarité, les brevets, le mentorat des doctorants et l'obtention des financements. En analysant 32,947 chercheurs en STEM observés entre 2006 et 2015, j'étudie l'effet d'être affilié à des universités qui ont postulé et qui ont reçu IDEX. De plus, j'étudie l'effet indirect sur les chercheurs d'universités non candidates à IDEX qui collaborent avec des chercheurs d'universités primées. En utilisant une approche difference-in-differences, je trouve que le fait de postuler et d'être récompensé par IDEX élargit les collaborations des chercheurs. IDEX stimule particulièrement les collaborations avec d'autres universités françaises et les collaborations internationales. Je trouve des effets indirects positifs de IDEX sur les collaborations des chercheurs dans les universités non candidates.

Dans le troisième chapitre, je compare l'efficacité de deux modèles de financement de la recherche : le financement en bloc et le financement compétitif. Les gouvernements comptent aujourd'hui sur des subventions compétitives pour allouer des fonds à la recherche, plutôt que sur le financement en bloc traditionnel. La littérature n'a pas encore répondu à la question suivante : la recherche financée par subventions compétitives a-t-elle un plus grand impact de celle financée en bloc ? Dans le contexte français, je compare l'impact de 6,441 articles scientifiques résultant de subventions compétitives avec 6,441 articles similaires résultant d'un financement en bloc. Je m'appuie sur les sections de remerciements des articles pour récupérer les informations de financement et sur les données de citation pour évaluer l'impact scientifique. J'applique une procédure d'appariement probabiliste pour la comparaison. Je constate que les articles bénéficiant des subventions compétitives obtiennent plus de citations sur le long terme que celles soutenues par un financement en bloc, alors que la différence n'est pas significative à court terme.

Ma thèse offre des informations importantes aux décideurs politiques dans la conception de politiques scientifiques efficaces.

Mots-clés : Doctorants, Chercheurs, Productivité scientifique, Financement IDEX, Financement compétitif, Financement en bloc.

ACKNOWLEDGMENTS

I sincerely thank Michele Pezzoni and Fabiana Visentin for their continued support, their friendship, and for the good time spent during our fruitful collaboration. I am extremely grateful for this. I thank Patrick Musso and Mark J. McCabe for their precious advice. I thank Raffaele Miniaci for encouraging me to take this path. I thank all my colleagues, the professors, and the staff of GREDEG. You made me feel at home from the first time I entered the lab.

I thank my family for understanding my choices and always supporting me. I thank Aurora for always being with me, for being able to appreciate me for who I am, and for the lovely time she devotes to me. This work is dedicated to my nephew Gabriele. I hope you will understand why I have not been able to be near you these years. You took care of our family with your pure smile. If one day you read this work, I hope it will be a source of inspiration for you.

This thesis has been supported by the French government, through the UCA^{JEDI} and EUR DS4H Investments in the Future projects managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-0001 and ANR-17-EURE-0004. This thesis was also supported by ERASMUS+ funding for a research stay at the United Nations University, Maastricht Economic and social Research institute on Innovation and Technology, the Netherlands.

CONTENTS

GENERAL INTRODUCTION

The role that policymakers attribute to science and the way they interact with it through the design of science policies have changed over time. The idea of the relevance of scientific research for economic growth was strengthened in the second half of the 20th century, after the world conflicts, evidenced by the creation of the National Science Foundation in 1950 in the US to provide public support for science.

Science has gradually acquired a central role as a determinant of countries' economic development. It has increasingly been considered a driver of innovation operating through several mechanisms, such as the knowledge transfer and supply of human capital to industry, its contribution to R&D and to the development of new industrial products and processes, university patenting, academic scientists' industry engagement, and indirectly through knowledge spillovers in the system.

The increasing importance of science is evidenced by the introduction of the expression "knowledge economy", where researchers and universities assume a central role (Foray, 2004; Geuna and Rossi, 2015). The idea of a knowledge economy has prompted governments to design incentives to support science. On the one hand, governments intend to use science as a lever to foster countries' international competitiveness, and, on the other hand, they aim to generate the amount of research that maximizes social benefits. If left to the private sector, research is likely to be underproduced due to its nature of public good (Nelson, 1959). The socially optimal amount of science requires substantial resources to buy up-to-date equipment and support risky research having an uncertain nature. Besides these reasons, international university rankings have triggered public investments in science. Governments invest in science also to improve the countries' scientific reputation and excellence (Stephan, 2012).

Government incentives are mainly designed in the form of research funding. The European Union, which has historically relied on block funding, i.e., a steady stream of funding provided by the governments to universities on a formula bases, is moving in recent years towards a competitive allocation of financial resources, which echoes the model adopted historically in the US based on contractual funding to universities and researchers. This approach implies quasi-market incentives for universities and researchers that are asked to compete with each other for a pre-determined amount of money (Geuna, 2001; Stephan, 2012).

With the adoption of competitive mechanisms, EU governments aim, on the one hand, to stimulate research excellence by increasing the efficiency in the allocation of resources and, on the other hand, addressing research agendas of universities and researchers towards objectives of social and economic interests. These objectives relate to the modern societal challenges that governments have to face, such as climate change and digitalization. In line with the new competitive funding allocation mechanism, the EU established the European Research Council (ERC) in 2007, to "support excellent investigators and their research teams to pursue ground-breaking, high-gain/high-risk research" (ERC, 2017). The ERC relies on contractual funding and is based on competitive mechanisms to allocate resources. As an example of the influence on research agendas, in 2021, the European Research Area (ERA) presented the Communication "A new ERA for Research and Innovation" to set the EU research lines. ERA claims "strategically orient and prioritise R&I investments"[1] to achieve the green and digital transitions. Within the ERA, the EU has introduced the 'Horizon Europe 2030' competitive funding programs to grant the best researchers with the best research ideas.

Along with its own funding instruments, the EU encouraged reforms of the member states' higher education systems and the creation of national funding agencies. De Boer et al. (2017) have identified over 30 structural reforms in the European states' higher education systems since the 1990s. They found common features characterizing reforms. A typical way to reform the HE system is through vertical differentiation, i.e., the creation of performance differences between higher education institutions. With vertical differentiation reforms, governments aim to improve the excellence of a selected group of universities through incentives based on a competitive allocation of resources. The most common approach consists of an "Excellence initiative" stimulated by the distribution of competitive funding. Some examples are the Spanish *International Campus of Excellence* initiative launched in 2008, the German *Excellence Initiative* launched in 2006, the Danish *Investment Capital for University Research* initiative launched in 2009, and the French *Initiative D'Excellence* (IDEX) launched in 2011. The case of the French IDEX funding is relevant to this thesis. The French government launched IDEX in 2011 as part of the PIA national fiscal stimulus (*Plan d'Investissements pour l'Avenir*) in response to the economic crisis. IDEX is a competitive funding program with which the French government distributes resources to universities on a competitive basis to support them in reaching excellence at a global level.

Along with the reforms in higher education, national funding agencies have been created. They consist of agencies under the direct authority of a ministry, normally the ministry of higher education. Across the member states, they share the common objective of promoting research excellence. Their main goal is to select and fund high-quality research projects. After organizing competitive calls for research proposals, they award a certain number of projects based on a peer review selection process.

[1] https://era.gv.at/public/documents/4427/Item_5_Renewed_and_new_RD_investment_targets.pdf

Their resource allocation is influenced by the research lines promoted by European policies. Relevant to this thesis is the French national funding agency, l'*Agence Nationale de la Recherche* (ANR), created in 2005. ANR is also the operator for the IDEX program, with the main role of distributing funding and monitoring the projects.

The change in the EU rationale for supporting science, with the introduction of higher education reforms and competitive calls, has raised many concerns among scholars (Geuna, 2001; Stephan, 2012). It is still not clear to what extent moving towards competitive mechanisms to allocate resources with quasi-market incentives is beneficial for science. The market mechanisms introduced by the competitive model imply that resources are distributed based on a measure of performance that requires an evaluation process for universities and researchers. However, quantifying the effects of policy incentives in science is arduous. Scientific outputs are characterized by multifaced aspects involving education, research, and innovation. Moreover, the concentration of resources into the hands of a few awarded universities and researchers depending on their performance may aggravate inequalities in science. Due to the self-reinforcing mechanisms that occur in science (Merton, 1968), where current recognition depends on past success, a quasi-market allocation of resources is likely to contribute to inequalities by feeding the process of cumulative advantages.

Pressure on performance and self-reinforcing mechanisms are likely to affect especially young scholars. Small differences in productivity at the beginning of the scientific career lead to large inequalities among researchers over time. Furthermore, the situation for young scholars is aggravated by fierce competition in the job market after the Ph.D. In the last 20 years, the OECD countries almost doubled the number of graduate students while the number of high skills job positions did not increase at the same pace (Cyranoski et al., 2011). An efficient design of the Ph.D. training programs is needed to allow for an optimal allocation of resources that guarantees effective training to increase the employability of young scholars.

In addition to these concerns, recent studies have evidenced a possible problem of risk-averse behaviors of funding agencies (Franzoni et al., 2022). Agencies tend to avoid failure and promote safe research, i.e., research demonstrating low outcome uncertainty that can produce results in the immediate term. Doing so, the research funding model based on competitive mechanisms may disfavor novel and breakthrough research. Furthermore, other costs for researchers and universities may emerge due to the adoption of the competitive model. Researchers are required to spend additional time on applications and administrative tasks that distract them from the research activity. Universities face instead research constraints imposed by contractual funding that can harm their research potential, discourage long-term research projects, and lead to internal conflicts. Moreover, the time required to adapt to the new mechanism is likely to harm universities' knowledge production.

Recent literature is trying to find answers to all these concerns. But still, studies are not exhaustive and many questions remain unanswered. Concerning the impact of funding on science, recent studies focus on the effect of individual grants awarded to established researchers, neglecting funding incentives addressed to Ph.D. students or universities. The main reason for this literature gap lies in the difficulty of collecting data both for young scholars, due to the multifaced aspects of the environment where they are trained and begin their career, and for universities, due to the multi-nature of the universities' outputs. The result is that there is no current consensus on how financial incentives affect the performance of researchers and universities. Moreover, existing studies focus on selected samples of researchers, often from top-tier universities, specific disciplines or cohorts.

With this thesis, I contribute to filling these literature gaps offering insights to policymakers aiming to achieve an efficient allocation of resources in science. In doing so, I consider the role played by three key actors in the knowledge production process: Ph.D. students, researchers, and universities. All the empirical analyses in this dissertation are based on the French case. I study the French research landscape in the context of the recent reforms in the higher education system that introduced competitive mechanisms for resource allocation and contractual-oriented incentives for science, leading to the creation of the ANR funding agency and the launch of the IDEX funding program. Both ANR and IDEX are expected to influence the performance of Ph.D. students, researchers, and universities and the quantity and quality of their scientific outcomes. Specifically, I address the following research questions: What makes a productive Ph.D. student? What is the impact of the IDEX initiative of excellence funding on French researchers' outcomes? Is grant-funded research more impactful than block-funded research?

In the following outline, I briefly introduce each chapter describing the main objectives and results.

CHAPTER 1

The increased number of graduate students that almost doubled in the last 20 years in the OECD countries has not been compensated by the increase in the number of high skills job positions. This trend has determined a fierce competition among young scholars for job positions available after the Ph.D. Policymakers aiming at graduating students highly competitive in the job market urge to understand the determinants of effective training programs and find levers for designing policies that maximize students' productivity. In other words, they urge replying to the question: What makes a productive Ph.D. student?

Although other studies have contributed to understanding what makes a Ph.D. student productive, they show some limitations. Specifically, previous studies (i) rely on selected samples of students in specific universities, disciplines, or cohorts, (ii) overlook the influence of peers' characteristics, and (iii) do not consider the role of supervisor's competitive grants and university funding.

In my first chapter, I rely on a unique dataset covering the entire population of 77,143 STEM Ph.D. students who graduated between 2000 and 2014 from French universities. I investigate the effect of a broad range of characteristics of the social environment a Ph.D. student is exposed to on the student's scientific productivity during the Ph.D. period. I measure the student's productivity as her publication quantity, publication quality, and scientific network size.

I find that students in productive environments are more productive, according to almost all the productivity measures considered. Having a female supervisor is associated with higher student productivity, especially in engineering, the most male-centered discipline in my sample. Surprisingly, mentorship experience is associated with lower Ph.D. student productivity, while having a mid-career supervisor is associated with higher student productivity. Having a supervisor with a French or European competitive research grant is associated with a higher number of citations received by the student. However, being supervised by a researcher awarded competitive European grants negatively relates to the student's publication quantity and network size. Once controlled for university characteristics, the government university funding program for excellence (IDEX) does not relate to the student's productivity. Sharing the training experience with large groups of peers penalizes the student's productivity, most likely due to a decline in the quality of the mentorship activity in large groups. On the contrary, having freshman peers, peers who publish high-quality articles, and at least one female peer is positively associated with the student's productivity.

CHAPTER 2

In the second chapter, I assess the impact of the IDEX funding program on a broad range of university researchers' activities. The extant literature trying to quantify the impact of research funding is scant, and it mainly focuses on grants awarded to individual researchers. The extant few studies on universities do not provide a clear view of the extent to which government funding affects universities. Moreover, they differ in the outcomes analyzed and mainly focus on one specific outcome at a time, overlooking the multifaced aspects of universities' activities.

The IDEX funding program was launched by the French government in 2011. Analyzing a panel of 32,947 researchers in STEM disciplines observed between 2006 and 2015, I estimate the impact of IDEX funding on researchers from seventeen applicant universities, eight awarded universities, and nine non-applicant universities. I consider several aspects of researchers' activities that can be traced back to the goals of IDEX, such as publication productivity, research interdisciplinarity, collaboration networks, patenting, mentoring, and fundraising.

I contribute to the literature by analyzing for the first time the impact of university funding considering a broad set of university researchers' outcomes. Moreover, in addition to the effect of obtaining IDEX, I investigate the effect of applying for funding and the effect of funding spillovers, two aspects often neglected by the previous literature. I study the latter effect by estimating the indirect impact of IDEX on researchers affiliated with universities that did not apply for IDEX but who collaborate with researchers affiliated with universities awarded IDEX.

Using a difference-in-differences approach, I find that both being affiliated with a university that applies for IDEX and being affiliated with a university that is awarded IDEX affect researchers' outcomes, but the effect is limited to the researchers' network. Being awarded IDEX is particularly beneficial for enlarging the researchers' collaborations with researchers from other French universities and international collaborations. Moreover, I find positive spillovers of IDEX for researchers in universities that did not apply for funding but collaborated with researchers in universities awarded IDEX. I find heterogeneous results when investigating the impact of IDEX in different research fields, but consistent evidence of the impact of IDEX on researchers' networks.

CHAPTER 3

In the final chapter, I investigate the effects of different funding models used to support science. Specifically, I compare the impact of research produced with the support of competitive grant funding with the impact of the research resulting from block funding. EU policymakers are increasingly relying on the competitive model to allocate science funding.

Although scientific literature has already shown that funding is one of the levers for knowledge creation, the literature analyzing the effectiveness of different funding models is still scant. The question of whether the impact of knowledge produced through the competitive funding system differs from that produced under the block funding approach is still unanswered. Answering this question may shed light on the possible risk-averse behavior of funding agencies.

I contribute to the literature by comparing the impact of the research produced through competitive grants distributed by the French main funding agency, l'Agence Nationale de la recherche (ANR), with the impact of research resulting from institutional block funding. To do so, I rely on scientific articles' acknowledgments to identify grant- and block-funded articles published between 2009 and 2013. I overcome the problem of using selected publication samples or disciplines, present in the extant literature, by identifying the entire set of scientific articles resulting from grants awarded by the ANR agency. Thus, I implement a probabilistic matching procedure to compare the impact of 6,441 grant-funded articles with the impact of 6,441 similar block-funded articles. To assess the articles' impact, I count the yearly citations they receive in the short and long run.

The main finding is that articles resulting from competitive grants receive about 7% more citations than articles resulting from block funding in the long run. In the short run, the difference is not statistically significant. Moreover, I find that articles in Mathematics follow a different pattern: when supported by grant funding, they are less impactful in the short run, while there is no difference in the long run. These findings can be explained both by the ANR agency's effort to support breakthrough research and the benefit that the additional resources of ANR grants provide to researchers.

CHAPTER 1

What makes a productive Ph.D. student?*

*This chapter is co-authored with Michele Pezzoni and Fabiana Visentin and has been published in *Research Policy*, with the reference "Corsini, A., Pezzoni, M., and Visentin, F., 2022. What makes a productive Ph.D. student? *Research Policy*, Volume 51, Issue 10, December 2022, 104561". (https://doi.org/10.1016/j.respol.2022.104561)

"My supervisor has everything I was looking for in a mentor. She is young and ambitious, and she overcomes any inexperience with a thirst for sharing her knowledge. Choosing me as her first PhD while establishing her own research group, filled me with a sense of responsibility while giving me the freedom to create something that I consider my own."

(Testimonial by a second-year Ph.D. in Human Medicine)[2]

"Professor A's group has developed many multidisciplinary research frontiers. From his connections, I have the opportunities to work with excellent colleagues in the School of Medicine. The collaborative research experiences during my PhD study are beneficial for me to expand my expertise toolkit. All the group members in Professor A's lab are very productive and the atmosphere in the group has been very enjoyable. The size of the group is just right, and the group is very dynamic and collaborative."

(Testimonial by a graduate student in Electrical engineering)[3]

1.1 Introduction

In the last 20 years, the OECD countries almost doubled the number of graduate students, passing from 154,000 in 2000 to 276,800 in 2017 (OECD, 2019, 2013), while the number of high skills job positions did not increase at the same pace (Cyranoski et al., 2011; Sauermann and Roach, 2012). This trend has determined a fierce competition for job positions available after the Ph.D. (Freeman et al., 2001). A recent article in *Nature career news* surveying 317 early-career researchers seeking academic positions warned students who want to undertake an academic career that at least 15 job applications are needed to receive a single job offer (Fernandes et al., 2020; Notman and Woolston, 2020). The hyper-competition in the job market requires Ph.D. students to focus on the outcomes with high value for recruiters, who select candidates showing a solid publication profile and a rich scientific network (Alberts et al., 2014). Despite the call from the scientific community to give less weight to publication metrics in the selection decisions (Benedictus et al., 2016), the practice of publication and citation counting persists, and the norm for Ph.D. students is to publish their thesis chapters even before graduation (Black and Stephan, 2010; Brischoux and Angelier, 2015; Horta and Santos, 2016; Sauermann and Haeussler, 2017; van Dijk et al., 2014). In addition to the publication record, recruiters also value the candidate's scientific network (Heffernan, 2021) as a signal of the candidate's capacity to establish research collaborations in an era in which knowledge production is increasingly the result of a team effort (Katz and Martin, 1997; Wuchty et al., 2007). Therefore, students, directors of Ph.D. programs, and policymakers aiming at graduating students highly competitive on the job market urge to understand which working conditions are associated with students' high publication scores and large scientific networks. In other words, they urge replying to the question: what makes a Ph.D. student productive?

[2] https://www.findaphd.com/advice/blog/4554/the-best-thing-about-my-phd-supervisor-students-share-their-stories
[3] https://www.ese.wustl.edu/~nehorai/students/testimonials.html

In this paper, we address this question by analyzing the role that a broad range of characteristics of the social environment to which a Ph.D. student is exposed has on the student's scientific productivity during the Ph.D. period. Doing so, we provide three contributions to the extant literature.

First, extant studies show scattered empirical evidence on how factors such as university quality, supervisor's gender, supervisor's scientific network, student's nationality, students group specialization, and funding relate with students' productivity (Baruffaldi et al., 2016; Conti et al., 2014; Gaulé and Piacentini, 2018, 2013; Horta et al., 2018; Pezzoni et al., 2016; Rossello et al., 2020; Waldinger, 2010). Our paper encompasses in a unique analysis a comprehensive set of relevant biographic and academic characteristics of the supervisor and peers, the two most important actors with whom the student establishes relationships during the training period (Carayol and Matt, 2004; Shibayama et al., 2015; Stephan et al., 2007; Stephan and Levin, 1997).

Second, although the student-supervisor relationship has already received attention (Paglis et al., 2006; Platow, 2012; Sinclair et al., 2014), some relevant supervisor characteristics have been neglected by the past empirical literature, such as supervisors' mentorship experience and fundraising ability. Moreover, the empirical literature has often overlooked the influence of peers' characteristics during the student's training period (with some notable exceptions such as Broström, 2019). Nonetheless, students spend most of their time in labs, frequently interacting with their peers, making group dynamics fundamental for the student's learning process (Shibayama and Kobayashi, 2017). Our paper contributes to advance knowledge on how neglected supervisor characteristics and peers' characteristics are associated with the student's productivity.

Third, extant studies on students' productivity rely on selected samples of students affiliated to top-tier universities (Pezzoni et al., 2016), working in specific disciplines (Delamont and Atkinson, 2001; Gaulé and Piacentini, 2018), or graduating in specific years (Broström, 2019; Shibayama and Kobayashi, 2017). Using selected samples is sometimes highly desirable allowing for solid identification strategies (Waldinger, 2010). However, it comes at the cost of limited external validity of the results of the analyses due to the necessity of drawing conclusions for specific disciplines, universities, or historical periods. Our analysis overcomes this limitation by covering all French universities in all STEM fields over a long time span, including 15 cohorts of students.

The study most similar to ours is a paper by Broström (2019). Brostöm investigates how department conditions relate to Ph.D. students' early career success. He employs data on Swedish students and finds that they perform better in the early stages of their careers when trained in small teams and supervised by a professor with a solid academic profile. A key difference from our work is that Brostöm looks at a selected sample of surveyed students who graduated from one cohort and work in group, while we use data on the entire population of one country, including 15 cohorts of graduate students. Having the whole population, our study does not suffer the representativeness concerns of survey data, and the long-time span observed allows us to control for cohort effects. Another important difference from our work is that Brostöm investigates the relationship between the Ph.D. environment and postgraduation outcomes. Former Ph.D. students might have entered very different job contexts, with some students working in highly reputed universities after graduation while others quitting academia. The postgraduation environment might drive part of the identified effects. In contrast, our work bounds Ph.D. students' productivity during their training period, associating the social environment with outcomes strictly related to the training period. Our results are, therefore, informative of the effectiveness of Ph.D. programs.

Our results show that higher student productivity is associated with having a productive female supervisor. On the contrary, having a supervisor with long mentoring experience and a supervisor in early- or late-career phases is associated with lower student productivity. The supervisor's fundraising ability at the national and European level is associated with higher visibility of the Ph.D. student's work, as shown by the citations received by the student's Ph.D. publications. However, being supervised by a researcher awarded European grants negatively relates to the student's publication quantity and network size. Looking at peers' characteristics, we find that a high number of peers is associated with a lower student's productivity. Conversely, having freshman highly-cited peers is positively associated with the student's productivity. When we break down our analysis by field, i.e., Mathematics, Engineering, Physics, and Medicine-biology-chemistry, we find heterogeneous results across fields.

1.2 Social environment and productivity during the training period

Understanding how the social environment characteristics are associated with the productivity of Ph.D. students is under the spotlight in the current discussion within the scientific community (Chenevix-Trench, 2006; Lempriere, 2020). This discussion has become of primary importance due to the sharp rise in the number of Ph.D. holders in the last decades that have not corresponded to an equal rise in the number of research job positions (Cyranoski et al., 2011). The mismatch between the supply and demand of Ph.D. graduates has strengthened the competition for the few available positions (Brischoux and Angelier, 2015; Freeman et al., 2001; Mangematin, 2000; van Dijk et al., 2014). When asked, more than one-third of French Ph.D. students declare nowadays to be worried about their professional future (Pommier et al., 2022). In the hyper-competitive context created, Ph.D. graduates aiming at pursuing a research career are mainly evaluated on their publication outcomes during their training period and their collaboration network (Alberts et al., 2014; Heffernan, 2021;

Mangematin and Robin, 2003). Therefore, publication outcomes and collaboration networks have become fundamental assets in determining graduates' career success (Allison and Stewart, 1974; Long and McGinnis, 1985; Merton, 1968; Tenenbaum et al., 2001; Vale, 2015).

As in any other working context, students' achievements measured by publication outcomes and collaboration networks depend on the characteristics of the environment to which students are exposed. During the Ph.D. training, the relevant environment for academic training is the lab where a student works (Shibayama and Kobayashi, 2017). Within the lab, students develop social relationships with their supervisors and peers. The successful completion of the Ph.D. and the graduate program satisfaction depend on these relationships (Lovitts, 2001; Tompkins et al., 2016). Therefore, we expect supervisor and peers' characteristics to be associated with the student's productivity.

1.2.1 Supervisor's characteristics and student's productivity

Student's productivity largely depends on the success of the student-supervisor collaboration. As in any other collaboration relationship, collaborators' characteristics play a crucial role in determining the success of the collaboration. In the case of scientific collaborations, biographic characteristics and academic profile are essential elements to consider (Azoulay et al., 2010; Bercovitz and Feldman, 2011; Bozeman and Corley, 2004; Katz and Martin, 1997; Lee and Bozeman, 2005; Taylor and Greve, 2006). Even more so, in the student-supervisor collaboration where the supervisor's characteristics are expected to be crucial for the student's productivity due to mentorship and lab leadership role played by the supervisor (Delamont and Atkinson, 2001; Golde, 2005; Lee et al., 2007; Lempriere, 2020; Liénard et al., 2018; Ma et al., 2020; Mangematin and Robin, 2003; Pearson and Brew, 2002; Shibayama, 2019; Shibayama et al., 2015; Tenenbaum et al., 2001).

Looking at the supervisor's biographic characteristics, we consider gender, seniority, and mentorship experience. We expect female and male supervisors to have different mentorship approaches. Ethnographic studies investigating the lab routines have explored these behaviours. Surveying 185 students at the University of California, Tenenbaum, Crosby, and Gliner, (2001) find that male supervisors are less likely than their female counterparts to provide psychological help to the students decreasing students' level of satisfaction with the Ph.D. training experience. However, female and male supervisors offer equal "instrumental help," providing students the same technical knowledge needed to support their publication productivity. In another survey study in medicine, Luckhaupt et al. (2005) find that female supervisors perceive gender-related boundaries in collaborating with their students. In a field experiment involving the recruitment of students in lab manager positions, 127 professors evaluating students' resumes have been funded to favor male students (Moss-Racusin et al., 2012). Larger empirical studies have confirmed those differences. In a sample of 20,000 US Ph.D. graduates in chemistry, Gaule and Piacentini (2018) find that students pairing with a same-gender advisor are more productive than students working with an advisor of a different gender. Similar results have been found for South African Ph.D. students by Rossello et al., (2020), who show that female students working with male supervisors are less productive than male students. In the context of a leading US interdisciplinary university, Pezzoni et al. (2016) find that having a female supervisor increases Ph.D. students' productivity.

Another supervisors' biographic characteristic that is expected to affect students' productivity during the training period is the supervisors' seniority. A rational individual decreases the working time with seniority (Diamond, 1984; Levin and Stephan, 1991). In the case of scientists, we expect that they allocate their time differently across activities as seniority increases. Indeed, scientists have a high degree of autonomy in choosing the time to allocate to different activities such as fundraising, research, teaching, consulting, and administrative activities (Libaers, 2012; Sabatier et al., 2006). We expect that young supervisors aiming to boost their careers devote more time to fundraising, research, and mentoring activities. In contrast, senior supervisors are likely to dedicate more time to remunerative activities in the short term, such as consulting and administrative activities. Consequently, the less time spent in research and mentoring activities by a senior supervisor might negatively impact the support provided to her Ph.D. students, and ultimately on her students' productivity.

While seniority is expected to harm students' productivity, we expect a positive relationship between the supervisor's mentorship experience and student's productivity. Accumulated experience in supervising students develops different abilities, such as advising, tutoring, encouraging, providing a role model, and conveying to students technical and tacit knowledge (Broström, 2019; Overington, 1977). Therefore, the supervisor's mentoring skills are expected to evolve with experience and lead to better student training when the supervisor has a long mentoring history. This better training is expected to be associated with the higher productivity of the Ph.D. student during the Ph.D. period.

Looking at the supervisor's academic profile, we consider her publication record, scientific network, and fundraising abilities. Publications and citations received reflect the supervisor's academic status and scientific competencies. Ph.D. students supervised by highly productive scientists are expected to acquire practical knowledge on how to conduct successful research (Long and McGinnis, 1985; Sinclair et al., 2014). Indeed, the supervisor often becomes a model for the student who reproduces the same successful research methodologies, develops similar skills and competencies, and applies the same commitment to research enterprises (Paglis et al., 2006).

Mimicking a productive supervisor's successful behaviour is expected to increase the student's productivity during the Ph.D. period.

The dimension of supervisors' network is also expected to be associated with students' productivity. For example, students supervised by scientists in contact with many co-authors are expected to be more likely to spend visiting periods in other labs acquiring new competencies, be introduced to leading scientists in the discipline, and be exposed to different research approaches (Mangematin and Robin, 2003; Stephan, 2006). These networking opportunities are expected to positively impact students' productivity (Lee and Bozeman, 2005).

Besides publication and networking influence, supervisors are also fundamental in providing resources that contribute to the students' Ph.D. program completion. Scholars have focused on assessing the role played by different types of scholarships on students' productivity (Horta et al., 2018). However, modern labs have 'firm-like' characteristics (Etzkowitz, 2003), making their competitiveness and survival substantially dependent on the amount of funds the professor leading the lab can raise (Stephan, 2012). Supervisors' fundraising activity is essential to support students' conference participation, visiting periods in other research institutes, and access to up-to-date lab equipment. Therefore, the supervisor's abundance of research funding is expected to be positively related with the Ph.D. student's productivity during the training period.

1.2.2 Peers' characteristics and student's productivity

Our study considers the student's peers as the other students exposed to the same work environment, i.e., having the same supervisor as the focal student during the same training period (Conti et al., 2014; Luckhaupt et al., 2005).

Ph.D. students, like any other worker, interact with peers during their professional activities. These interactions might affect students' productivity in several ways. First, students feel the "peer pressure" of maintaining a level of productivity similar to that of their peers striving for scientific recognition from their supervisor and the scientific community (Stephan and Levin, 1992). Moreover, the comparison with productive peers triggers psychological mechanisms of social comparison, making the focal student adopting the same productive behaviours as her colleagues (Tartari et al., 2014). Finally, students learn by observing and interacting with their peers stimulating the generation of novel research ideas (Ayoubi et al., 2017; Cornelissen et al., 2017; Delamont and Atkinson, 2001). Although peer pressure and learning from peers mechanisms are expected to increase the student's productivity during the training period, coordination costs and competition dynamics might be detrimental to large groups' productivity (Broström, 2019).

The labour literature, both using observational and experimental data, is convergent in showing that having peer co-workers in the work environment positively affects productivity (Falk and Ichino, 2006). However, we expect the beneficial effect of having peers shrinking when the peers' number increases (Shibayama and Kobayashi, 2017). Indeed, the supervisor's time allocated to each student might reduce when the number of students increases, and the upsurge of competitive dynamics between peers might discourage students' collaboration (Conti et al., 2014).

Not only the mere presence of peers affects the focal student's productivity, but also peers' characteristics. Similar to the supervisor, we analyze peers' biographic and academic characteristics.

As biographic characteristics, we expect that both the gender and seniority of peers are associated with students' productivity. Previous studies have not reached convergent results on gender. Looking at undergraduate students, Dasgupta et al. (2015) find that group dynamics are not gender-neutral. For instance, female students' participation and self-confidence in group discussions are higher in female-majority groups. Looking at Ph.D. students, Pezzoni et al. (2016) found that, although student and supervisor's gender matters, the gender composition of the lab is not associated with the Ph.D. student's productivity. Regarding the peers' seniority, having more senior peers with greater knowledge stocks is expected to enhance knowledge transfer toward the focal student (Ayoubi et al., 2017; Delamont and Atkinson, 2001), and it might increase students' productivity. However, more senior peers might be in a phase of their Ph.D. when ideas are already settled, and interacting with other students might be less fruitful.

As peers' academic characteristics, we consider peers' publication and citation productivity. Previous literature has shown that peers' productivity positively affects individuals' productivity for low-skilled jobs such as supermarket workers and fruit-pickers (Bandiera et al., 2009; Mas and Moretti, 2009). For high skilled jobs, such as scientific research, results are not convergent. While Azoulay et al. (2010) show a decrease in the scientific productivity of team members when the team "star scientist" dies, Waldinger (2012) finds no effect of losing a brilliant peer. Although these not convergent results, in the Ph.D. students' context, we expect that highly productive peers will benefit the focal student's productivity through the three mechanisms described above: "peer pressure" adoption of productive behaviours inspired by peers through the mechanism of social comparison, and enhanced probability of acquiring knowledge from productive peers.

The mechanism of social comparison might also play a role in encouraging the expansion of the focal student's network. Although we have argued that students mainly rely on their supervisor's network to create their collaboration network, students surrounded by peers who invest energies in developing their co-authorship network during conference participation and visiting periods probably will tend to mimic the same behavior. Therefore, we expect the student's network size to be larger when peers have a larger network.

1.3 The French population of STEM Ph.D. students

Our empirical setting is represented by the entire population of STEM Ph.D. students of one European country, France. The excellence of France in STEM fields is proved by the worldwide recognition gained by its scholars and top-tier research institutions. Looking at the absolute number of Nobel Prize winners, 39 French scientists obtained the highest recognition in Chemistry, Medicine, and Physics. A French elite institute, the *École Normale Supérieure in Paris*, is ranked first together with the *California Institute of Technology* by the proportion of alumni who obtained the prize. Marie Curie, the first woman who obtained a Nobel Prize and the only woman awarded twice, received her training mainly in Paris, where she established her lab. France does exceptionally well also in Mathematics, being one of the top-5 countries for the number of Fields medals.

In training scientists, France has a well-structured doctoral offer. Ph.D. scholarships are sponsored by universities, laboratories, the State, or private companies. Students are supported by scholarships that usually last three years (Pommier et al., 2022)[4]. Students' hiring contracts are relatively standard, and almost all students are hired as full-time professional researchers for the entire duration of their Ph.D. (Mangematin, 2000). A centralized system standardizes doctoral program regulations, but each university has margins of flexibility in organizing courses and lab activities. Usually, programs show field heterogeneity. For instance, Ph.D. students in natural and technological sciences work full time in research labs with their colleagues, while in the other disciplines, students' work does not require a daily presence in labs. During their first year, Ph.D. students are often asked to attend core classes in theory and methodology and additional skill classes such as "writing scientific papers". In later years, a considerable amount of a student's Ph.D. time is devoted to writing the thesis, a document of about 200 pages where the student proves her research abilities. The prevalent thesis format has evolved over time, from producing a coherent monography on a specific subject to the current standard of producing a collection of three independent research articles. This change is in line with the attempt to encourage young scholars to publish their Ph.D. research work in scientific journals to facilitate their future careers. The final thesis importance is evident from the fact that French researchers often interchange the expression "being enrolled in a Ph.D. program" with "*faire une these*" (the English equivalent of "writing a thesis"). Candidates need to be paired

[4] In 2021, 97% of French Ph.D. students in Science and Technology fields benefitted from specific funding to support their Ph.D. training (Pommier et al., 2022).

with a thesis supervisor who accepts to guide them to access the doctoral program. The practice of writing a thesis under the guidance of a supervisor assisted by a co-supervisor is allowed.

1.4 Data sources

To construct our study sample, we gather data from multiple sources. The first is the French repository of *Electronic Doctoral Theses*. By special permission, we obtained access to the whole universe of STEM thesis records collected by the *Agence Bibliographique de l'Enseignement Supérieur* (ABES) that is managing the repository since 1985. For each thesis record, we have information about author, abstract, university of graduation, defense date, supervisor's name, co-supervisor's name (if any), and field of study. As fields, we distinguished theses in Mathematics, Engineering, Physics, and Medicine-biology-chemistry[5]. Unfortunately, the records do not report the student's year of entry into the Ph.D. program; thus, we approximate it assuming that each student started the program three years before her thesis defense year. According to the national statistics for STEM fields, the most frequent duration of the Ph.D. training in France is four years, three years plus the thesis defense year.[6] Hence, we set the student's *entry year* into the Ph.D. program in year *t-3*, and we define the *Ph.D. training period* as the period ranging from *t-3* to *t*, where *t* is the defense year.

Our information on the students' and supervisors' gender results from a multiple-iteration matching strategy (Gaulé and Piacentini, 2018; OECD, 2012). First, we match the students' given names with the official French gender-name dataset.[7] Then, for the non-matched names, we repeated the matching exercise with the *U.S. Census Bureau* gender-name dataset and the WIPO gender-name dataset[8], respectively.

We retrieve students' and supervisors' publication records from Elsevier's SCOPUS database. We match the ABES list of students with the SCOPUS authors affiliated to French institutions using students' names and surnames as key matching criteria[9]. Similarly, we match supervisors' names and surnames with the SCOPUS authors.

---

[5] We also used a fine-grained distinction of fields based either on the Scopus field classification of supervisors' publications or on a manual attribution of the theses. The results of the fine-grained regression exercises are consistent with the ones presented in the main text. Results are available upon request.
[6] The Ph.D. duration is consistent with the duration of the scholarships. We double checked this statistic by querying a subset of universities' administration.
[7] Website: https://www.data.gouv.fr/fr/datasets/liste-de-prenoms/
[8] Website: https://www.wipo.int/publications/en/details.jsp?id=4125
[9] We dropped from the initial list of students provided by ABES students with homonymous names. Having two or more students with the same full name in our original list of Ph.D. thesis authors would make it difficult to disentangle their identity and correctly assign bibliometric information. Therefore, we decided to drop the homonyms from our original list of Ph.D. thesis authors.

We gather information on funding at the national and the European level. At the national level, we use the complete list of individual grants awarded by the *Agence Nationale de la Recherche* (ANR), the French national funding agency. Outside France, we consider the funding programs at the European level. We use the list of individual grants, *Horizon 2020* (H2020) and *Framework Programmes* (FP), awarded by the European Commission and collected in the CORDIS dataset. We match supervisors with principal investigators using their names and surnames.

To reconstruct the quality of the Ph.D. students' graduation department, we rely on the *QS university ranking*[10]. The QS university ranking provides detailed information on the universities' academic reputation at the department level and allows us to flag the top departments in each field. For instance, *Université de Paris* is in the top-20 percent of universities in Mathematics in France, but not in Engineering. We integrate the information from the QS ranking with bibliometric information concerning the university affiliates. We construct an appropriate bibliometric dataset of the publications and publications' authors for all the French university departments. To create this dataset, we manually match the names of the French universities (and their variants) with the SCOPUS affiliations' names. As an additional proxy for the department quality, we identify the French universities that in 2011 benefitted from the *Initiative D'Excellence* (IDEX) funding provided by the French Government to a selected group of French higher education institutions. The IDEX funding program was launched in 2011 by the French Government within a national fiscal stimulus and awarded to eight universities[11] striving to become competitors of worldwide top-ranked universities.

To create our study sample, we link all the information retrieved from the data sources listed above in a unique original dataset. Doing so, we joined student, supervisor, and department information. In addition, we refined our study sample excluding students showing productivity indicators too high to be credible[12]. Overall, the excluded students represent less than 10% of our initial sample from the ABES list of student names. After this cleaning exercise, we obtained a study sample of 77,143 Ph.D. students who graduated from French universities between 2000 and 2014.

[10] Website: https://www.topuniversities.com

[11] The 8 awarded universities are: Université d'Aix-Marseille, Université de Bordeaux, Université Paris Saclay, PSL Paris Sciences et Lettres, Sorbonne Université, Sorbonne-Paris-Cité, Université de Strasbourg, Université de Toulouse.

[12] We excluded students with more than 20 publications, more than 100 citations received per paper, and more than 200 co-authors during the Ph.D. period. We excluded also students for which their supervisors reported more than 100 publications and more than 500 co-authors during the five years preceding the student enrollment.

1.5 Econometric methodology

To estimate how the Ph.D. student's social environment characteristics relate to her productivity, we estimate the coefficients of the model presented in Equation 1 using *Ordinary Least Squares* (OLS). As represented by subscript *i*, the analysis is at the student level.

$$Student's\ productivity_i = \beta_0 + (Supervisor's\ characteristics_i)'\beta_1 + (\ Peers'characteristics_i)'\beta_2 + (Controls_i)'\beta_3 + \varepsilon_i$$

Equation 1

The left-hand side variable *Student's productivity* in Equation 1 takes, in turn, the value of the student's publication quantity, quality, and the size of the scientific network. We measure the publication quantity by counting the number of peer-reviewed papers published by the student (*Publications*) and the publication quality by counting the number of yearly citations received on average by the student's papers (*Average citations*). We proxy the student's research network size as the number of the student's distinct co-authors (*Co-authors*). The three productivity variables are calculated during the Ph.D. training period, i.e., from *t-3* to *t*, with the addition of one year after the thesis defense to account for possible time lags in the publication process (Powell, 2016). In other words, we calculate the productivity outcomes in the period ranging between *t-3* and *t+1*, where *t* is the thesis defense year.

The vectors *Supervisor's characteristics* and *Peers' characteristics* define the Ph.D. student's social environment. *Controls* is a vector including the student's characteristics and the characteristics of the department where the student is enrolled. Finally, ε is the idiosyncratic error term. Our interest is to estimate the vectors of coefficients β_1 and β_2 that relate supervisor and peers' characteristics with the student's productivity.

A concern in estimating these coefficients regards a potential endogeneity issue. Although we include in our regression a large set of time-variant and time-invariant characteristics identified by the previous literature as factors affecting the student's productivity, the lack of proxies for the student's intrinsic ability might bias our estimates. Indeed, an omitted variable problem might arise if the unobserved ability correlates with explained and explanatory variables. For instance, students with higher research ability might be at the same time more productive and more likely to be supervised by scientists with better academic credentials. However, previous studies have shown that this endogeneity problem is mitigated by the supervisor's difficulty in assessing the student's research ability when the student is at the beginning of her academic career (Mangematin, 2000). The asymmetry of information during the student's selection process makes it unlikely to observe a correlation between students' intrinsic ability and supervisors' quality. Belavy et al. (2020) show in an empirical study on 324 Ph.D. students that variables usually used as proxies for the students' ability, such as previous academic achievements and training grades, are uncorrelated with the students' Ph.D. productivity. Along the same line, anecdotal evidence shows that standardized tests often considered for Ph.D. enrollment, e.g., GRE scores in the US, do not fully reflect the student's future academic ability (Aristizábal, 2021). Although previous literature excludes a strong correlation between the student's academic ability and the supervisor's quality, in Appendix E, we implement a robustness check to respond to the potential endogeneity concern. Specifically, we replicate the estimations of Equation 1 adding a proxy that controls for the ability of the student during her high school period. We flag students with exceptional ability by calculating a dummy variable equal to one if the student has participated in a selective contest during high school (Agarwal and Gaule, 2020). We consider three well-known contests: the *International Mathematical Olympiad* (IMO), *Les Olympiades Nationales de Mathématiques* (the national French Mathematical Olympiad), and *le Kangourou des mathématiques* (a French national mathematical contest). We find that including a proxy for the student's ability does not affect the estimated coefficients of the variables in the *Supervisor's characteristics* and *Peers' characteristics* vectors, showing that our results are unlikely to be affected by an endogeneity problem.

1.5.1 Supervisor's characteristics

We consider the supervisor's biographic and academic characteristics. As for the biographic characteristics, we include a dummy variable *Female supervisor* that equals one if the supervisor is a female scientist, zero otherwise. Expecting that the attention dedicated to a Ph.D. student varies along the supervisor's career, we calculate the *Supervisor's seniority* measured as the years elapsed between the supervisor's first publication and the student's entry year into the Ph.D. program. To capture possible nonlinear effects of seniority, we include a squared term of the variable *Supervisor's seniority*. Also, the mentorship experience of the supervisor might affect the productivity of her Ph.D. students. Therefore, we calculate the variable *Mentorship experience* as the cumulated number of students mentored by the supervisor who have successfully defended their thesis until the focal student's entry year into the Ph.D.[13]

Concerning the supervisor's academic characteristics, we calculate two variables proxying the supervisor's publication quantity and quality in the five years preceding the entry of her student into the Ph.D. program, i.e., from *t-8* to *t-4*, where *t* is the student's defense year. We decided to measure the supervisor's publication quantity and quality during the five years preceding the student

[13] We retrieve data on supervisors' mentoring career starting from 1980.

enrollment (and not during the student training period) because it is a common practice that the student and her supervisor co-sign articles during the student's training period. In the case of co-signed articles, it is impossible to disentangle supervisor and student's productivity. We define the variable *Supervisor's publications* as the number of supervisor's publications in peer-reviewed journals over the five years preceding the student's entry into the Ph.D. program. Then, for the same period, we calculate the average number of yearly citations received by the supervisor's articles (*Average citations*). To proxy for the supervisor's scientific network size, we reconstruct her co-authorship network. We define the variable *Supervisor's co-authors* as the number of distinct co-authors that the supervisor has in the five years preceding the student's entry into the Ph.D. program. Finally, to proxy for the supervisor fundraising ability, we calculate a dummy variable *ANR grant* that equals one if the supervisor is the principal investigator of an ANR grant in at least one year of the student's training period. Similarly, we define a dummy variable *EU grant* that equals one if the supervisor is the principal investigator of at least one EU grant during the student's training period.

1.5.2 Peers' characteristics

Ph.D. students might spend their Ph.D. training period alone if their Ph.D. period does not overlap with the Ph.D. period of other students. In the opposite case, they might share the Ph.D. experience with other peer students. To distinguish these two cases, we calculate the dummy variable *With peers* that takes value one if the focal student spends at least one year of her training period with at least another student having the same supervisor, zero otherwise. Then, we calculate the variable *N. peers* as the average yearly number of students with whom the focal student shares the training experience. Students start their Ph.D. training in different moments, and cohorts of students can overlap only partially. To calculate the variable *N. peers*, we first calculate the yearly number of peers in each of the four years of the focal student's training period; then, we average the four values. For instance, if the focal student spends the first three years alone and her supervisor recruits another student during her last Ph.D. year, the variable *N. peers* equals 0.25 (0.25=(0+0+0+1)/4).

To characterize the relationships between the student and her peers, we calculate variables proxying for the peers' biographic and academic characteristics. Concerning the biographic characteristics, we calculate the dummy variable *At least one female peer* that equals one if at least one peer during the focal student's training period is a female student, zero otherwise. We also calculate the peers' average seniority as the average number of years spent by the peers in their Ph.D. program (*Average peers' seniority*). Also, in this case, peers might have training periods that only partially overlap with that of the focal student. Thus, as the first step of the peers' seniority variable construction, we calculate the average peer seniority in each year of the 4-years of the focal student's training period. If the focal student has no peers in one year, we assign the value zero to the average yearly seniority. Then, we obtain the *Average peers' seniority* variable averaging the four yearly values. For instance, if the focal student has only one peer during the first two years of her training period, it means that the peer defended her thesis during the focal student's second year of Ph.D. Thus, we consider the peer's seniority values for the first two years of the focal student's training equal to 3 and 4. The variable *Average peers' seniority* equals 1.75 (1.75=(3+4+0+0)/4) for the focal student.

Concerning the academic characteristics, we calculate the peers' number of publications per year (*Peers' publications*). This variable is calculated following a two-step procedure. In the first step, we count the number of articles published by the peers in each of the four years of the focal student's training period. In case the focal student has no peers in one year, we assign the value zero to the yearly number of articles published. Then, we obtain the *Peers' publications* by averaging the four values. For instance, if the focal student has two peers who publish one article each[14] during the first year of her training period, the value of *Peers' publications* equals 0.5 (0.5=(2+0+0+0)/4). Applying the same two-step procedure as for the *Peers' publications*, we calculate the variable *Peers' average citations* proxying for the quality of peers' work and the variable *Peers' co-authors* proxying for the peers' network size.

1.5.3 Other controls

To mitigate a potential bias of our estimated coefficients, we control for the department and student's characteristics. We define a department as the pair university-field. For instance, *Université de Paris* counts four departments: *Université de Paris*-Mathematics, *Université de Paris*-Engineering, *Université de Paris*-Physics, and *Université de Paris*-Medicine-biology-chemistry.

To control for department quality, we retrieve the university reputation ranking from the QS World University ranking.[15] We create a dummy *French Top-20* that equals one if the department is among the 20% of departments with the highest academic reputation in a specific field in France. As an additional proxy for the department quality, we calculate the average citation-weighted publication productivity per department affiliate (*Citation-weighted publications per affiliate*). To calculate this latter variable, we consider the department affiliates' average productivity during the five years preceding the student's entry into the Ph.D. program. Specifically, we identify the department affiliates' publications during the five years preceding the student enrollment; then, we weigh each

---

[14] In case of joint publications between two or more peers of the same focal Ph.D. student, we count the publication once.
[15] https://www.topuniversities.com/university-rankings. We gather the ranking information in 2020, however university ranking has minor variation over the years when considering top-universities. The advantage of using the QS World University ranking is the availability of a ranking that is detailed by subject area.

publication by the citations received each year. Finally, we calculate the average number of affiliates' citation-weighted publications for each department. We also calculate the variable *IDEX* as a third control for the department quality. This variable is a dummy that equals one after 2011 if the student's department was selected to be awarded the IDEX national investment program funding.

To control the department size, we calculate the variable *Department size* counting the number of scientists affiliated to the department for at least one year during the five years preceding the student's entry into the Ph.D. program. We rescale the number of affiliates dividing by 100, meaning that each unit increase of the variable *Department size* corresponds to 100 additional department affiliates[16].

Along with the department size, the size of the Ph.D. program might also play a role. Larger Ph.D. programs might be better organized and provide students with a better and more productive training experience. We calculate the number of Ph.D. students enrolled in the same focal student's Ph.D. program for each of the four years of her training period. Then, we calculate the variable *N. of Ph.D. students in the program* averaging the four yearly values.

Finally, we control for the characteristics of the Ph.D. student. Specifically, we control for the gender of the student with a dummy variable *Female student* that equals one for female students, zero otherwise.[17] We consider the student's possibility of having a thesis co-supervisor defining the dummy *Co-supervision* that takes value one in the presence of a co-supervisor, zero otherwise. We also add four dummy variables, *Mathematics*, *Engineering*, *Physics,* and *Medicine-biology-chemistry* controlling for the heterogeneity across the thesis research fields. Finally, we add a set of dummy variables for the students' *Entry year* into the doctoral program to account for the Ph.D. cohort effect.

1.5.4 Descriptive statistics

Table 1 lists all the variables included in our analysis with a short description. Table 2 reports the descriptive statistics for the variables calculated on our sample of 77,143 Ph.D. students. When classified by field, 15% of the students are in Mathematics, 18% in Physics, 21% in Engineering, 45% in Medicine, Biology, and Chemistry. Students publish on average 2.37 peer-reviewed articles during their training period. 68% percent of students publish at least one article during the Ph.D. period. The average students' collaboration network includes 8.93 distinct co-authors.

The average supervisor has a stock of 13.59 peer-reviewed articles and a seniority of 11.49 years of career when her student enrolls in the Ph.D. program. At the time of the student enrollment, the

[16] In an alternative model specification, we include department fixed effects. Our results are unchanged and available upon request.

[17] We do not have information about the age of the Ph.D. students, however in France students tend to enroll in the Ph.D. program soon after their master studies, thus we do not expect much age heterogeneity among students.

average supervisor counts 3.08 successfully supervised Ph.D. students over her career. For the gender composition, 39% of Ph.D. students are women, while this percentage reduces to 21% when looking at the supervisors. Only 6% of the students have a supervisor who is the principal investigator of an ANR national grant during the Ph.D. training period, and only 2% of the students have a supervisor who is the principal investigator of a EU grant.

Looking at the focal Ph.D. student's peers, 80% of the students have at least one peer during the training period, and, on average, they are in contact with 1.76 peers per year. During the training period, the focal student's peers publish on average 0.81 papers per year.

Table A1, in Appendix A, reports the variable correlation matrix.

Table 1. List of variables used in the analysis.

	Variable description
Dependent variables	
Student's productivity	
Publications	Ph.D. student's number of papers published between t-3 and t+1*
Average citations	Average yearly citations received by the student's papers published between t-3 and t+1
Co-authors	Number of distinct co-authors of the student between t-3 and t+1
Independent variables	
Supervisor characteristics	
Female supervisor	Dummy variable that equals one if the supervisor is a female scientist
Supervisor's seniority	Number of years elapsed from the first supervisor's publication to t-3
Mentorship experience	Cumulated number of Ph.D. students successfully supervised until t-3
Supervisor's publications	Supervisor's number of papers published between t-8 and t-4
Supervisor's average citations	Average yearly citations received by the supervisor's articles published between t-8 and t-4
Supervisor's co-authors	Supervisor's number of distinct co-authors between t-8 and t-4
ANR grant	Dummy variable that equals one if the supervisor is the principal investigator of an ANR grant between t-3 and t
EU grant	Dummy variable that equals one if the supervisor is the principal investigator of a EU grant between t-3 and t
Peer characteristics	
With peers	Dummy variable that equals one if the student has at least one peer between t-3 and t
N. peers	Average number of the student's peers per year between t-3 and t
At least one female peer	Dummy variable that equals one if at least one student's peer is a female student between t-3 and t
Average peers' seniority	Average yearly seniority in the Ph.D. program of the student's peers
Peers' publications	Average number of peers' publications per year between t-3 and t
Peers' average citations	Average yearly citations received by the peers' articles between t-3 and t
Peers' co-authors	Peers' average number of distinct co-authors per year between t-3 and t
Other controls	
French Top-20	Dummy variable that equals one if the student's department is among the 20% departments with the highest academic reputation score in France according to the QS ranking
Citation-weighted publications per affiliate	Average department affiliate's citation-weighted publication productivity between t-8 and t-4
IDEX	Dummy variable that equals one if t is greater or equal to 2011 and the student is enrolled in a university awarded IDEX funding
Department size [100 affiliates]	Total number of scientists affiliated to the student's department between t-8 and t-4
N. of Ph.D. students in the program	Average number of Ph.D. students per year enrolled in the focal student's Ph.D. program between t-3 and t
Female student	Dummy variable that equals one if the Ph.D. student is female
Co-supervision	Dummy variable that equals one in the presence of a co-supervisor
Mathematics	Dummy variable that equals one if the Ph.D. dissertation is in Mathematics
Engineering	Dummy variable that equals one if the Ph.D. dissertation is in Engineering
Physics	Dummy variable that equals one if the Ph.D. dissertation is in Physics
Medicine-biology-chemistry	Dummy variable that equals one if the Ph.D. dissertation is in Medicine, Biology, or Chemistry
Entry year	The student's entry year into the Ph.D. program, i.e., t-3

NOTE: *t is the Ph.D. thesis defence year; t-3 is the entry year of the student into the Ph.D. program; the four years ranging from t-3 to t define the Ph.D. training period; the five years ranging from t-8 to t-4 are the years preceding the student's entry into the Ph.D. program.

Table 2. Descriptive statistics for our sample of 77,143 Ph.D. students.

77,143 Ph.D. students	Mean	SD	Min	Max
Dependent variables				
Ph.D. student				
Publications	2.37	2.99	0.00	20.00
Average citations	2.11	3.51	0.00	98.14
Co-authors	8.93	15.37	0.00	200.00
Independent variables				
Supervisor characteristics				
Female supervisor	0.21	0.41	0.00	1.00
Supervisor's seniority	11.49	5.24	0.00	21.00
Mentorship experience	3.08	6.22	0.00	184.00
Supervisor's publications	13.59	14.31	0.00	100.00
Supervisor's average citations	2.36	3.03	0.00	127.87
Supervisor's co-authors	37.28	50.82	0.00	499.00
ANR grant	0.06	0.25	0.00	1.00
EU grant	0.02	0.16	0.00	1.00
Peer characteristics				
With peers	0.80	0.40	0.00	1.00
N. peers	1.76	2.14	0.00	30.00*
At least one female peer	0.52	0.50	0.00	1.00
Average peers' seniority	1.61	1.04	0.00	3.56
Peers' publications	0.81	1.76	0.00	41.00
Peers' average citations	2.71	8.11	0.00	353.15
Peers' co-authors	4.21	10.28	0.00	190.75
Other controls				
French Top-20	0.39	0.49	0.00	1.00
Citation-weighted publications per affiliate	7.37	4.43	0.38	35.05
IDEX	0.18	0.38	0.00	1.00
Department size [100 affiliates]	29.25	30.28	0.04	114.46
N. of Ph.D. students in the program	1042.07	800.94	1.00	2973.00
Female student	0.39	0.49	0.00	1.00
Co-supervision	0.31	0.46	0.00	1.00
Mathematics	0.15	0.36	0.00	1.00
Engineering	0.21	0.41	0.00	1.00
Physics	0.18	0.39	0.00	1.00
Medicine-biology-chemistry	0.45	0.50	0.00	1.00
Entry year	2005.12	4.20	1997.00	2011.00

NOTE: *Although the maximum number of peers might look high, we checked the case of the student with 30 peers during the training period. The student was supervised by a researcher in Physics, having yearly 30(+1) Ph.D. students during the focal student's training period.

1.6 Results

Table 3 reports the OLS estimates of the model described in Equation 1.

Table 3. Regression results. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Supervisor characteristics			
Female supervisor	-0.0051	0.074**	0.31**
	(0.025)	(0.030)	(0.13)
Supervisor's seniority	0.037***	0.0071	0.11***
	(0.0072)	(0.0085)	(0.036)
Supervisor's seniority^2	-0.0019***	-0.00096**	-0.0067***
	(0.00034)	(0.00040)	(0.0017)
Mentorship experience	-0.018***	-0.0072***	-0.037***
	(0.0019)	(0.0023)	(0.0097)
Supervisor's publications	0.027***	0.0070***	-0.10***
	(0.0012)	(0.0015)	(0.0062)
Supervisor's average citations	0.031***	0.20***	0.21***
	(0.0036)	(0.0043)	(0.018)
Supervisor's co-authors	0.0028***	0.0014***	0.091***
	(0.00034)	(0.00040)	(0.0017)
ANR grant	0.0048	0.54***	0.22
	(0.043)	(0.050)	(0.21)
EU grant	-0.19***	0.33***	-1.28***
	(0.065)	(0.077)	(0.33)
Peer characteristics			
With peers	0.13***	0.24***	0.25
	(0.041)	(0.048)	(0.21)
N. peers	-0.12***	-0.042***	-0.39***
	(0.0071)	(0.0083)	(0.036)
At least one female peer	-0.028	0.073**	0.21*
	(0.025)	(0.030)	(0.13)
Average peers' seniority	-0.14***	-0.13***	-0.63***
	(0.017)	(0.020)	(0.086)
Peers' publications	0.13***	-0.15***	-0.64***
	(0.014)	(0.016)	(0.070)
Peers' average citations	0.0065***	0.056***	0.049***
	(0.0020)	(0.0024)	(0.010)
Peers' co-authors	0.0029	0.0017	0.21***
	(0.0023)	(0.0027)	(0.011)
Other controls			
French Top-20	-0.0082	0.068**	-0.36***
	(0.023)	(0.028)	(0.12)
Citation-weighted publications per affiliate	0.012**	0.026***	0.14***
	(0.0057)	(0.0067)	(0.029)
IDEX	-0.056	0.031	-0.032
	(0.036)	(0.042)	(0.18)
Department size [100 affiliates]	0.00081	0.0014**	0.013***
	(0.00057)	(0.00067)	(0.0029)
N. of Ph.D. students in the program	0.000092***	0.00023***	0.00038***
	(0.000016)	(0.000019)	(0.000079)
Female student	-0.64***	-0.19***	-1.84***
	(0.021)	(0.025)	(0.11)
Co-supervision	-0.066***	-0.042	-0.66***
	(0.023)	(0.027)	(0.12)
Engineering	0.18***	0.40***	0.99***
	(0.035)	(0.041)	(0.18)
Physics	0.77***	0.57***	2.46***
	(0.055)	(0.065)	(0.28)

Medicine-biology-chemistry	1.54***	1.39***	6.45***
	(0.043)	(0.050)	(0.22)
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	1.23***	0.23***	3.83***
	(0.074)	(0.087)	(0.37)
Observations	77,143	77,143	77,143
R-squared	0.140	0.128	0.174

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. Standard errors are reported in parentheses. In an additional exercise, we calculate the p-values applying a multiple-inference adjustment to correct possible erroneous inferences due to the high number of hypotheses tested. Specifically, we calculate the p-values applying the Romano-Wolf multiple hypothesis correction (Romano and Wolf, 2005a, 2005b; Romano and Wolf, 2016). The statistical significance of the coefficients remains consistent with the main results, with only two notable exceptions. Specifically, the coefficient of the variable *At least on female peer* loses its statistical significance (at standard significance levels) in the regression explaining the number of student's C*o-authors* and the coefficient of the variable *Department size [1000 affiliates]* loses its statistical significance in the regression explaining the *Average citations* received by the student's work. The exercise is available upon request.

Looking at the impact of the biographic characteristics of the supervisor on the student's productivity, we find that having a *Female supervisor* is not associated with the number of papers published by the student. On the contrary, having a female supervisor is associated with a higher number of citations (+0.074 yearly citations per paper) and a larger collaboration network (+0.31 co-authors). These two variations are statistically significant and have economic relevance, corresponding to the 3.5%[18] of the sample average student's citations and 3.5% of the sample average student's co-authors. Regarding the *Supervisor's seniority,* we find an inverted U-shape relationship between the supervisor's seniority and the three student's outcomes considered. The maximum impact of seniority on the student's publication productivity, citations, and network size is for a mid-career supervisor, i.e., when the supervisor has 9.74[19], 3.70, and 8.21 years of seniority, respectively.

We find that the supervisor's *Mentorship experience* is negatively associated with the student's productivity: a student mentored by an experienced supervisor shows fewer papers published, citations received, and a smaller collaboration network. Increasing by one standard deviation, the *Mentorship experience* is associated with 0.11 fewer papers[20], 0.045 fewer citations, and 0.23 fewer co-authors. To further investigate this result, in Appendix F, we search for non-linear relationships between *Mentorship experience* and student's productivity. Specifically, we calculate a set of dummy variables identifying different levels of experience. Figure 1 reports the graphical representation of the marginal effects of these dummy variables as estimated in Table F1. Consistently with the results in Table 3, Figure 1 shows that the higher the supervisor's experience, the lower the student's

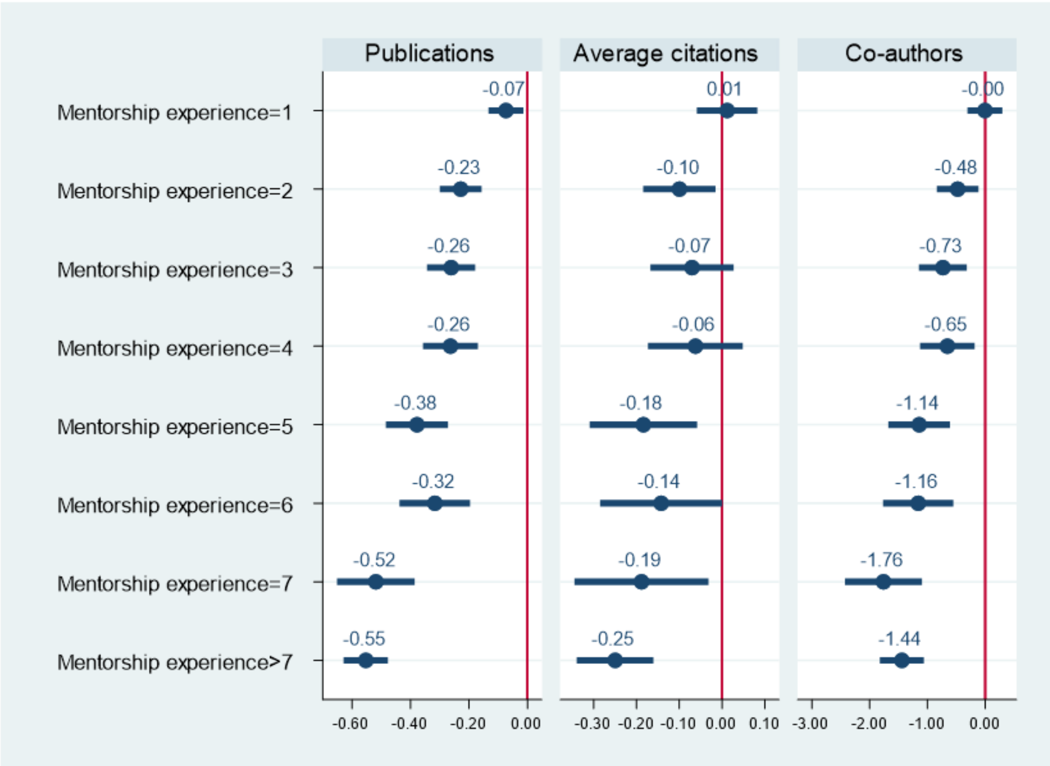
[18] This percentage is calculated dividing the variation of the student's *Average citations* associated to having a *Female supervisor* by the average value of *Average citations* in the sample, reported in Table 2 (2.11).

[19] The seniority corresponding to the maximum marginal effect on publication productivity is calculated using the coefficients estimated in Column 1 of Table 3, and applying the following calculation -0.037/(2*-0.0019).

[20] The value -0.11 is obtained by multiplying the coefficient of *Mentorship experience* estimated in Table 3, Column 1, and the standard deviation of *Mentorship experience* reported in Table 2 (-0.11=-0.018*6.22).

productivity. Focusing on the 11.7% of students supervised by researchers with a high *Mentorship experience*, i.e., researchers who supervised more than seven students before the current one, those students show 0.55 fewer publications (23.2% of the sample average), 0.25 fewer citations per paper (11.8% of the sample average), and 1.44 fewer co-authors (16.1% of the sample average) than the students supervised by mentors with no experience. This result contrasts with our expectation that being mentored by an experienced supervisor is positively associated with student's productivity. We interpret our finding as the supervisors' tendency to be more supportive to students when they are at their first experience as thesis directors[21].

Figure 1. *Mentorship experience* marginal effects on student's productivity outcomes.



NOTE: The figure reports the marginal effects estimated for the set of 8 dummy variables calculated in Appendix F and used in the regression exercises reported in Table F1. The variable *Mentorship experience=1*, takes value one if the supervisor has mentored only one Ph.D. student who graduated before the focal student enrollment. The variable equals one for 15.1% of the supervisors. Similarly, we calculate *Mentorship experience=2* (9.7%), *Mentorship experience=3* (7.0%), *Mentorship experience=4* (5.1%), *Mentorship experience=5* (3.9%), *Mentorship experience=6* (2.9%), *Mentorship experience=7* (2.4%), and *Mentorship experience>7* (11.7%). The reference case, represented by the vertical

[21] Interestingly, supervisor seniority is weakly correlated with the mentorship experience. This shows that, in our sample, we might observe supervisors in the early stages of their careers who accumulated a considerable mentorship experience and, *vice versa*, senior supervisors with no Ph.D. students. Moreover, in additional empirical analyses, we investigate the publication productivity distribution and the presence of CNRS affiliated researchers among supervisors with no previous supervision experience. We find that the publication productivity distribution for supervisors with no previous mentorship experience largely overlaps the productivity distribution of researchers with experience, meaning that there are high-quality researchers with notable publication records also among the supervisors without supervision experience. Similarly, we find that the proportion of researchers affiliated to CNRS is similar for supervisors with no supervision experience and supervisors with experience.

line centered in zero, is when the supervisor has *No mentorship experience* (42.1%). Bars represent 95% confidence intervals.

Looking at the supervisor's academic characteristics, supervisor's productivity measured by *Supervisor's publications, average citations,* and *co-authors*, is associated with higher student's productivity. Specifically, increasing the supervisor's publication by one standard deviation is associated with 0.39[22] additional student publications (16.3% of the sample average[23]) and 0.10 additional citations (4.75% of the sample average). Similar to *Supervisor's publications*, both the *Supervisor's average citations* and *co-authors* are associated with positive outcomes for the student along all the three dimensions considered. Increasing by one standard deviation the *Supervisor's average citations* is associated with 0.09 additional articles (3.96% of the sample average), 0.61 additional citations (28.72% of the sample average), and 0.64 additional co-authors (7.13% of the sample average). Increasing by one standard deviation the *Supervisor's co-authors* is associated with 0.14 additional articles (6.00% of the sample average), 0.07 additional citations (3.37% of the sample average), and 4.62 additional co-authors (51.79% of the sample average). The only exception to all these positive correlations is the relationship between the supervisor's number of publications and the student's network size: increasing the supervisor's publication by one standard deviation is associated with 1.43 fewer co-authors (16.02% of the sample average). This negative association might be explained by the fact that when students work with highly productive supervisors, they have fewer incentives to enlarge their network outside the lab. Despite this latter negative association, our results show a positive relationship between the supervisor's academic characteristics and the productivity of the Ph.D. student.

Considering the supervisor's fundraising ability, when the supervisor is the principal investigator of a French ANR grant, the student's work receives 0.54 additional yearly citations per paper, which corresponds to 25.59% of the students' citation average in our study sample. Similarly, having a supervisor awarded a European grant is associated with an increase of 0.33 citations received by the student's work (15.64% of the citation average). In contrast, having a supervisor awarded a European grant is associated with 0.19 fewer publications (8.02% of the publication average) and 1.28 fewer co-authors (14.33% of the co-author average). These negative correlations might be explained by the additional time spent by the supervisor managing the EU grant. Indeed, EU grants are large international projects funded by the European Commission, and supervisors need to invest a relevant

[22] This value is obtained by multiplying the standard deviation of the variable *Supervisor's publications* 14.31 (Table 2) by the coefficient 0.027 of *Supervisor's publications* in Table 3, Column 1.
[23] This percentage is calculated dividing the variation of the student's *Publications* associated to one standard deviation increase of *Supervisor's publications* by the sample average value of *Publications* reported in Table 2 (2.37).

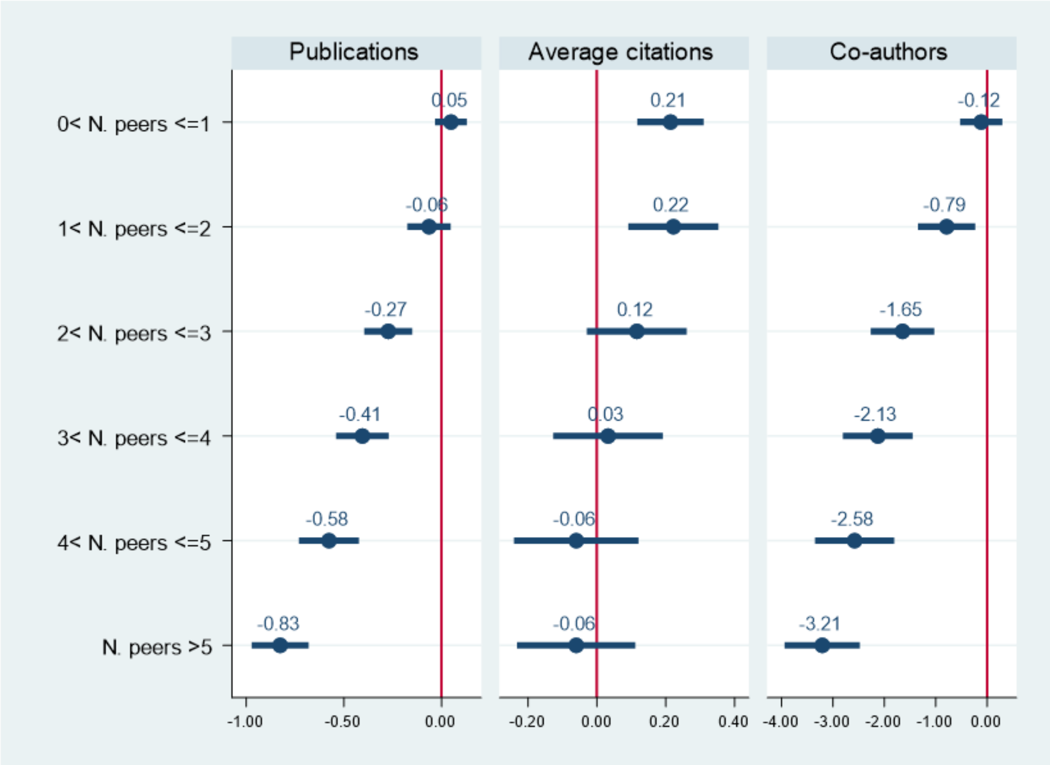
amount of time in managing them. This time is probably subtracted from mentoring students. Although we observe some differences between ANR national grants and European grants, our results converge in showing that the availability of supervisor's funds is positively associated with the quality of the student's productivity.

Looking at the peers' effect, we find a positive association between the dummy variable *With peers* and the Ph.D. student's productivity. However, this variable has to be always interpreted jointly with the variable *N. peers*, since when the dummy variable *With peers* equals one, the variable *N. peers* takes positive values. For instance, we find that the overall effect of having one peer in every year of the Ph.D. period is associated with 0.20 (=0.24-0.042*1) additional citations (9.4% of the sample average), and we do not observe any statistical significance[24] of having one peer for the publication quantity and co-authorship network size. Although having one peer is associated with benefits to citations, further increasing the number of peers is associated with a decrease in all dimensions of the student's productivity, namely 0.12 fewer publications, 0.042 fewer citations, and 0.39 fewer co-authors for each additional peer[25]. These three values correspond to 5.06% of the publication average, 2.00% of the citation average, and 4.37% of the co-author average in the study sample. This empirical evidence shows that the larger the number of peers, the lower the student's productivity. Therefore, sharing the training experience with large groups of peers penalizes students' productivity, showing that the quality of the mentoring activity declines if the supervisor has many students. This decline might be related to the lack of time devoted by the supervisor to each student. Moreover, this result is particularly relevant because it suggests an optimal number of peers associated with the student's productivity. In Table F2, Appendix F, we dig into these findings to identify possible non-linear relationships between the variable *N. peers* and the student's productivity. Specifically, we calculate six dummy variables, one for each unit increase in the value of the variable *N. peers*. The alternative model specification reported in Table F2 confirms the main results reported in Table 3: having up to one peer in each year of the Ph.D. period is associated with a higher number of citations received by the doctoral student's work. On the contrary, an increase in the number of peers is associated with a decrease in all three student's productivity outcomes. Figure 2 shows the marginal effects associated with an increasing number of peers in the student's environment on student's productivity outcomes.

[24] To test for the statistical significance of the linear combination of the coefficients of the variables *With peers* and *N. peers*, we conducted an F-test on the null hypothesis that $\beta_{With\ peers} + \beta_{N.of\ peers} * 1 = 0$.

[25] As a further robustness check, we run a regression selecting the subsample of 61,696 students with at least one peer. Results are consistent with those reported in Table 3.

Figure 2. ***N. peers*** **marginal effects on student's productivity outcomes.**



NOTE: The figure reports the marginal effects estimated for the 6 dummy variables calculated in Appendix F and used in the regression exercises reported in Table F2. The variable *0< N. peers <=1*, takes value one if the student has between 0 (excluded) and 1 (included) peers per year during the Ph.D. period. The variable equals one for 29.9% of the students. Similarly, we calculate *1< N. peers <=2* (20.7% of the students), *2< N. peers <=3* (12.2%), *3< N. peers <=4* (6.9%), *4< N. peers <=5* (3.8%), and *N. peers >5* (6.4%). The reference case, represented by the vertical line centered in zero, is when the focal student has *No peers* (20.0%). Bars represent 95% confidence intervals.

Conditional on having at least one peer, peers' biographic characteristics matter. Having *At least one female peer* student during the Ph.D. period is positively associated with both the focal Ph.D. student's citations received and network size, but not with the number of publications. The increase in the student's citations and co-authors equals 0.073 citations (3.46% of the sample average) and 0.21 co-authors (2.35% of the sample average). Increasing the variable *Average peers' seniority* by one standard deviation is associated with a lower focal Ph.D. student's productivity along all the dimensions considered, namely -0.15 publications (6.14% of the sample average), -0.14 yearly citations (6.41% of the sample average), and -0.66 co-authors (7.34% of the sample average). These results lead us to conclude that peers' gender positively correlates with the student's productivity, while peers' seniority negatively correlates with the student's productivity.

Regarding the peers' academic characteristics, an increase in the number of *Peers' publications* by one standard deviation is associated with fewer citations and fewer co-authors: -0.26 citations (12.51% of the sample average) and -1.13 co-authors (12.61% of the sample average). On the contrary, an increase in *Peers' publications* is associated with 0.23 additional articles published by the focal student (9.65% of the sample average). An increase of one standard deviation of the *Peers' average citations* is associated with an overall productivity boost for the focal student: +0.05 publications (2.22% of the sample average), +0.45 citations (21.52% of the sample average), and +0.40 co-authors (4.45% of the sample average). The increase of *Peers' co-authors* by one standard deviation benefits only the focal student's network size being associated with 2.16 additional co-authors (24.17% of the co-author sample average). In the light of these results, we conclude that peers' academic characteristics show mixed effects on the focal student's productivity. We can interpret these results on peers' productivity in the light of the "peer pressure" mechanism leading the student to maintain a productivity level similar to her peers. Specifically, when peers increase their publication quantity, the focal student feels the pressure to increase her outcomes in terms of quantity at the disadvantage of quality and collaboration aspects, consistently with the coefficients of the variable *Peers' publications* in the regression exercises. Differently, competition on quality between peers increases all the dimensions of scientific productivity considered, consistently with the coefficients of the variable *Peers' average citations* in the regression exercises.

For the controls, the quality of the department as measured by the variable *Citation-weighted publications per affiliate* is positively associated with all the students' productivity outcomes. On the contrary, when we measure department quality according to the variable *French Top-20*, we find that being affiliated to a top-20 reputed department positively relates to the student's citations while negatively relates to her network size. Finally, *French Top-20* is not significantly related to the number of articles published by the student. Doing a Ph.D. in a university benefitting from an *IDEX* award does not significantly correlate with the student's productivity outcomes.

The size of the department and the size of the Ph.D. student program do matter. The department size positively relates to the student's yearly citations and co-authors. Larger departments are more likely to generate internal collaborations between affiliates or attract a greater number of external collaborators. Similarly, an increase in the size of the Ph.D. program (*N. of Ph.D. students in the program*) is positively associated with all the Ph.D. student's productivity dimensions. Larger Ph.D. programs might be better structured and organized, benefitting students' productivity.

Considering the Ph.D. student characteristics, we find a significant gender gap between female and male students. Female students are less productive than their male counterparts across all the three outcomes investigated (-0.64 publications, -0.19 yearly citations, and -1.84 co-authors).[26] Moreover, the presence of a co-supervisor is associated with a decrease of the student's productivity.

[26] We have estimated an econometric model where we interacted the student gender with the supervisor gender. We found non-significant effects of the interaction terms. We do not report interactions in our main model specification.

Looking at the set of dummy variables identifying the fields of study, we observe productivity heterogeneity across fields. This latter result is expected since different fields are characterized by heterogeneous norms, rules, and working conditions affecting students' productivity. Following the idea that field heterogeneity matters, Section 1.7 explores the possibility of field-specific effects of our regressors by estimating the coefficients of Equation 1 for students in Mathematics, Engineering, Physics, and Medicine-biology-chemistry separately.

1.7 Further analyses

1.7.1 Exploring heterogeneity across fields

We leverage on our large data sample of students representing all the STEM fields to explore cross-field heterogeneity. Table 4 shows some structural differences across fields. On average, students in Mathematics are the least productive, with 1.12 papers published during the training period, 0.88 average yearly citations received, and a network composed of 2.59 distinct co-authors. On the contrary, Ph.D. students enrolled in Medicine-biology-chemistry are the most productive. They show an average productivity of 3.22 publications, 2.96 yearly citations received, and a large network of 13.39 co-authors. Table B1, in Appendix B, reports the descriptive statistics of the complete set of explanatory variables by field.

Table 4. Ph.D. students' productivity by field.

Dependent Variables	Engineering	Mathematics	Medicine-Biology-Chemistry	Physics
Publications	1.41	1.12	3.22	2.41
Average citations	1.27	0.88	2.96	1.97
Co-authors	4.00	2.59	13.39	8.79
Observations	16,519	11,450	35,038	14,136

Table 5 reports the estimations of the coefficients of Equation 1 by field. Looking at the supervisors' biographic characteristics, differently from the regressions presented in Table 3, the relationship between supervisor's seniority and student's productivity is not statistically significant in Engineering and Physics. The supervisor's mentorship experience shows the same negative association with all the student's outcomes across fields: the greater the number of students previously mentored by the supervisor, the lower the student's productivity outcomes. Having a female supervisor relates positively to students' productivity in Engineering, while the effect is limited in the other fields. Specifically, having a female supervisor in Engineering is associated with 0.25 additional publications, 0.29 additional yearly citations received, and 0.79 additional co-authors. This result is particularly interesting due to the specificities of Engineering if compared with other disciplines. Indeed, female supervisors in engineering are rarer (only 14% of the supervisors are female scientists) than in other disciplines (17% in Mathematics, 28% in Medicine-biology-chemistry, and 17% in Physics) (Hunt, 2010). Moreover, the few female supervisors observed in Engineering, if compared to their male counterparts, are more productive than female supervisors in other disciplines[27]. We interpret these facts as the result of a selection process that leads only women with outstanding scientific competencies to overcome all the obstacles to reach a professorship position in a male-dominated discipline such as Engineering. These female supervisors' outstanding competencies are beneficial for the supervised students who show higher productivity.

When looking at the supervisors' academic characteristics, having a strong publication profile has a positive relationship with all the Ph.D. students' productivity outcomes across fields. The only exception is the negative relationship between the supervisor's number of publications and the student's network size in Mathematics, Medicine-biology-chemistry, and Physics. The number of citations received by the supervisor's publications has a positive relationship with all the student's productivity outcomes across fields. When we consider the supervisor's scientific network, the correlation between the supervisor's number of co-authors and all the Ph.D. student's productivity outcomes is positive in Medicine-biology-chemistry, while it is limited to the student's network in the other fields.

Results in Table 5 show that being mentored by a supervisor who benefited from an ANR grant is positively associated with all the Ph.D. students' productivity outcomes in Engineering and Physics. When we consider European grants, we find that they are positively associated with students' citations in Physics and Medicine-biology-chemistry. This latter result might be explained by the high student visibility gained in these fields due to the collaboration with research teams in other European countries promoted by the international nature of European grants.

In all fields, the increase in the number of peers is associated with decreased student's productivity, with the only exception of the increase in citations received in Mathematics. Peers' seniority is associated with a productivity decrease of the focal student in Medicine-biology-chemistry and Physics, while it shows no correlation with productivity in Mathematics and a slightly negative correlation in Engineering. Having at least one female peer is associated with scattered productivity benefits across disciplines, except for Physics. In Engineering, having a female peer relates to an increase in the publication score and network size, in mathematics with an increase in the network size, in Medicine-biology-chemistry with an increase in the citations received.

[27] Looking at the publication score of female supervisors in engineering at the time of the students' enrollment, we find that their publication productivity is 77% of their male counterparts. In Mathematics is 69%, in Medicine-biology-chemistry is 65%, and in Physics is 69%.

Peers’ academic characteristics show mixed effects on students’ productivity outcomes. Interestingly, the peers’ network size is particularly favorable for the student’s productivity in Mathematics and Medicine-biology-chemistry, while the peers’ average citations benefit the student’s productivity in Medicine-biology-chemistry and Physics. The peers’ publication productivity is positively associated with the focal student’s publication productivity in Engineering, Medicine-biology-chemistry, and Physics.

Concerning the control variables, consistently with Waldinger’s study (2010) on mathematicians, we show a positive influence of the department's prestige on Ph.D. students’ productivity in Mathematics. However, we show also that this result does not hold for students in Engineering and Medicine-biology-chemistry. This finding highlights the importance of covering multiple fields when assessing the determinants of students’ productivity.

Table 5. Regression results, by field. OLS estimates.

	Engineering			Mathematics			Medicine-biology-chemistry			Physics		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Publications	Average citations	Co-authors	Publications	Average citations	Co-authors	Publications	Average citations	Co-authors	Publications	Average citations	Co-authors
Supervisor characteristics												
Female supervisor	0.25***	0.29***	0.79***	-0.098**	-0.028	-0.13	-0.050	0.033	0.11	-0.028	0.048	0.73**
	(0.049)	(0.059)	(0.20)	(0.048)	(0.069)	(0.21)	(0.039)	(0.046)	(0.20)	(0.065)	(0.072)	(0.37)
Supervisor's seniority	0.010	-0.00093	0.046	0.029***	0.012	0.036	0.027**	-0.038**	0.14**	0.016	0.0094	0.051
	(0.013)	(0.015)	(0.052)	(0.011)	(0.016)	(0.049)	(0.014)	(0.016)	(0.069)	(0.016)	(0.018)	(0.090)
Supervisor's seniority^2	-0.00024	-0.00029	0.00042	-0.00100*	-0.00040	0.00028	-0.0022***	0.00030	-0.011***	-0.00084	-0.00073	-0.0060
	(0.00060)	(0.00071)	(0.0025)	(0.00057)	(0.00081)	(0.0025)	(0.00063)	(0.00074)	(0.0032)	(0.00078)	(0.00085)	(0.0043)
Mentorship experience	-0.013***	-0.011***	-0.033***	-0.0024	-0.010**	-0.0067	-0.035***	-0.0070*	-0.11***	-0.035***	-0.021***	-0.084**
	(0.0028)	(0.0034)	(0.012)	(0.0031)	(0.0045)	(0.014)	(0.0034)	(0.0040)	(0.017)	(0.0062)	(0.0068)	(0.035)
Supervisor's publications	0.034***	0.024***	0.039***	0.033***	0.012**	-0.069***	0.023***	0.0021	-0.12***	0.041***	0.024***	-0.040***
	(0.0026)	(0.0031)	(0.011)	(0.0033)	(0.0047)	(0.014)	(0.0020)	(0.0024)	(0.010)	(0.0026)	(0.0029)	(0.015)
Supervisor's average citations	0.019**	0.12***	0.018	0.020***	0.083***	0.078***	0.024***	0.27***	0.27***	0.055***	0.22***	0.21***
	(0.0077)	(0.0093)	(0.032)	(0.0051)	(0.0072)	(0.022)	(0.0061)	(0.0072)	(0.031)	(0.0092)	(0.010)	(0.052)
Supervisor's co-authors	-0.0044***	-0.0042***	0.015***	-0.0026**	0.00041	0.064***	0.0067***	0.0027***	0.11***	-0.0038***	-0.0024***	0.063***
	(0.00086)	(0.0010)	(0.0036)	(0.0010)	(0.0015)	(0.0044)	(0.00054)	(0.00064)	(0.0028)	(0.00064)	(0.00071)	(0.0036)
ANR grant	0.26***	0.42***	0.78**	0.14	0.10	1.35***	-0.16**	0.60***	-0.84**	0.53***	0.49***	2.37***
	(0.083)	(0.100)	(0.35)	(0.090)	(0.13)	(0.39)	(0.065)	(0.076)	(0.33)	(0.11)	(0.12)	(0.61)
EU grant	-0.012	-0.040	0.18	-0.35**	-0.21	-2.01***	-0.38***	0.33***	-1.46***	0.20	0.57***	-1.07
	(0.12)	(0.15)	(0.52)	(0.15)	(0.21)	(0.64)	(0.10)	(0.12)	(0.53)	(0.14)	(0.15)	(0.78)
Peer characteristics												
With peers	0.12	-0.060	0.093	-0.049	0.085	-0.32	0.16**	0.31***	0.33	0.35***	0.34***	1.14**
	(0.081)	(0.097)	(0.34)	(0.077)	(0.11)	(0.33)	(0.067)	(0.079)	(0.34)	(0.093)	(0.10)	(0.52)
N. peers	-0.071***	-0.014	-0.22***	-0.048***	0.027*	-0.044	-0.27***	-0.13***	-1.01***	-0.14***	-0.048**	-0.53***
	(0.0100)	(0.012)	(0.041)	(0.010)	(0.015)	(0.044)	(0.015)	(0.018)	(0.077)	(0.021)	(0.023)	(0.12)
At least one female peer	0.093**	0.051	0.49***	0.032	-0.060	0.32*	-0.033	0.17***	0.25	-0.17***	-0.042	-0.022
	(0.039)	(0.047)	(0.16)	(0.043)	(0.061)	(0.18)	(0.046)	(0.054)	(0.23)	(0.061)	(0.067)	(0.34)
Average peers' seniority	-0.087***	-0.012	-0.16	-0.027	-0.045	0.030	-0.10***	-0.14***	-0.61***	-0.22***	-0.19***	-1.02***
	(0.031)	(0.037)	(0.13)	(0.031)	(0.044)	(0.13)	(0.029)	(0.034)	(0.15)	(0.042)	(0.046)	(0.23)
Peers' publications	0.12***	0.023	0.094	0.027	-0.079**	-0.47***	0.17***	-0.26***	-0.91***	0.23***	-0.051	-0.34
	(0.022)	(0.026)	(0.090)	(0.026)	(0.038)	(0.11)	(0.024)	(0.028)	(0.12)	(0.038)	(0.041)	(0.21)
Peers' average citations	-0.0048	0.018***	-0.028**	0.0012	0.0017	0.0027	0.011***	0.077***	0.11***	0.017***	0.075***	-0.0053
	(0.0033)	(0.0040)	(0.014)	(0.0038)	(0.0055)	(0.017)	(0.0032)	(0.0038)	(0.016)	(0.0059)	(0.0064)	(0.033)
Peers' co-authors	-0.0028	-0.0098**	0.055***	0.0092*	0.019***	0.14***	0.0085**	0.0088**	0.30***	-0.020***	-0.016**	0.16***
	(0.0036)	(0.0043)	(0.015)	(0.0048)	(0.0069)	(0.021)	(0.0038)	(0.0044)	(0.019)	(0.0058)	(0.0063)	(0.032)
Other controls												
French Top-20	-0.12**	0.050	-0.60***	-0.065	0.023	-0.21	-0.082**	0.052	-0.36*	0.18**	0.061	-0.25
	(0.050)	(0.060)	(0.21)	(0.047)	(0.067)	(0.20)	(0.037)	(0.043)	(0.19)	(0.079)	(0.086)	(0.44)
Citation-weighted publications per affiliate	0.035	0.00029	0.11	0.042**	0.064**	0.15**	-0.035*	0.017	-0.084	0.0044	0.021*	0.064
	(0.027)	(0.033)	(0.11)	(0.018)	(0.025)	(0.076)	(0.018)	(0.022)	(0.093)	(0.010)	(0.011)	(0.057)
IDEX	-0.12**	-0.00096	-0.61**	0.082	-0.063	-0.0055	-0.092	-0.033	0.25	0.13	0.18*	0.47
	(0.059)	(0.070)	(0.24)	(0.063)	(0.090)	(0.27)	(0.067)	(0.078)	(0.34)	(0.087)	(0.096)	(0.49)
Department size [100 affiliates]	-0.0016	0.014***	-0.024*	0.00086	0.027***	-0.0051	0.0049***	-0.00011	0.022***	0.0060***	-0.0031	0.040***
	(0.0033)	(0.0040)	(0.014)	(0.0054)	(0.0077)	(0.023)	(0.0011)	(0.0012)	(0.0054)	(0.0017)	(0.0019)	(0.0096)
N. of Ph.D. students in the program	0.00013***	0.000022	0.00066***	0.00013***	0.000082**	0.00030***	-0.000076***	0.00028***	-0.00022	0.00019***	0.00038***	0.0011***
	(0.000028)	(0.000033)	(0.00012)	(0.000026)	(0.000036)	(0.00011)	(0.000029)	(0.000034)	(0.00015)	(0.000043)	(0.000047)	(0.00024)
Female student	-0.33***	-0.15***	-0.75***	-0.29***	-0.17***	-0.44**	-0.84***	-0.21***	-2.63***	-0.64***	-0.22***	-1.83***
	(0.039)	(0.047)	(0.16)	(0.040)	(0.058)	(0.17)	(0.034)	(0.040)	(0.17)	(0.052)	(0.057)	(0.29)
Co-supervision	0.073**	0.094**	0.21	0.035	0.12**	0.11	-0.23***	-0.22***	-1.74***	0.11**	0.14**	0.32
	(0.037)	(0.044)	(0.15)	(0.040)	(0.058)	(0.17)	(0.041)	(0.048)	(0.21)	(0.054)	(0.060)	(0.30)
Entry year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.85***	0.26*	1.89***	0.97***	0.12	1.52***	3.73***	1.87***	15.3***	1.76***	0.52**	6.22***
	(0.13)	(0.15)	(0.52)	(0.12)	(0.16)	(0.50)	(0.22)	(0.26)	(1.11)	(0.21)	(0.23)	(1.19)
Observations	16,519	16,519	16,519	11,450	11,450	11,450	35,038	35,038	35,038	14,136	14,136	14,136
R-squared	0.042	0.038	0.032	0.045	0.029	0.052	0.087	0.101	0.142	0.087	0.110	0.079

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. Standard errors are reported in parentheses. In an additional exercise, we calculate the p-values applying a multiple-inference adjustment to correct possible erroneous inferences due to the high number of hypotheses tested. We rely on the Romano-Wolf multiple hypothesis correction (Romano and Wolf, 2005a, 2005b; Romano and Wolf, 2016). The statistical significance of the coefficients remains almost unchanged (at standard significance levels) across disciplines. The only notable exception is the coefficient of the variable number of *Peers' co-authors*, which loses its statistical significance in several regressions explaining student's productivity in Engineering, Mathematics, and Medicine-biology-chemistry. The exercise is available upon request.

1.7.2 Considering different types of publication outcomes

We construct our students' productivity measures considering all the publications produced by the student during her training period. However, those publications might result from different research activities. In particular, some publications might result from the joint work between the student and her supervisor, while others from autonomous work or form collaborations with other scientists. Furthermore, some publications might result from the core thesis work, while others might result from other research lines unrelated to the thesis.

This section presents two robustness checks to investigate how the environmental factors relate to these different types of students' publications. First, we select only publications listing among the authors, both the student's and her supervisor's name. Doing so, we identify the publications that result from the close collaboration between the student and her supervisor. Second, we isolate the publications deriving from the student's thesis work. To do that, we use a text analysis algorithm to compare the thesis and publications' content and select only the student's publications with similar content to her thesis manuscript.

On average, students co-author with their supervisors 1.76 publications, which corresponds to 74% of the publications attributed to the students in our main analysis. Student-supervisor coauthored publications receive on average 1.97 yearly citations and list 6.92 co-authors (see Table C1 of Appendix C). Re-estimating in Table C2 the models presented in Table 3 considering only student-supervisor coauthored publications, we find results consistent with Table 3 with a few exceptions. For instance, from Table C2, we observe that the coefficient of the variable *ANR grant* turns positive in the regression explaining publication quantity and student's coauthors. Indeed, having a supervisor awarded an ANR grant is associated with 0.16 additional student-supervisor publications and 0.71 additional co-authors.

In Appendix D, Table D1 reports the descriptive statistics of the three dependent variables calculated attributing to the student only publications similar to the thesis manuscript. To measure the similarity between the publications authored by the student and the content of her thesis manuscript, we rely on a text analysis algorithm[28] that compares publication abstracts with thesis abstracts (Mikolov et al., 2013). On average, students have 1.38 publications similar to their thesis manuscript, which corresponds to 58% of the publications attributed on average to the students in our main analysis. These publications receive 1.41 yearly citations and list 5.37 co-authors. Table D2 reports the regression estimates of Equation 1 using the three dependent variables considering only publications similar to the thesis. The regression results are consistent with our main analysis reported

[28] Appendix D provides details on the text analysis algorithm.

in Table 3, with a few exceptions. Like when looking only at the co-authored publications with the supervisor, also in the case of publications similar to the thesis, having a supervisor who is the principal investigator of an ANR grant positively correlates with student's publication quantity and co-authors. We interpret these results as the consequence of the pressure to publish experienced by ANR recipients who have to deliver publication outcomes as results of their project funded by the ANR agency. Therefore, supervisors with ANR grants tend to involve students in their projects, asking them to develop a thesis on ANR project topics and co-authoring with them. This involvement leads students to have a higher number of publications similar to the thesis and co-authored with the supervisor.

Interestingly, we observe also that the relationship between the supervisor's seniority and the Ph.D. student's productivity turns into a U-shaped relationship when we limit the analysis to publications similar to the thesis topics. This result differs from the inverted U-shaped relationship observed in the main analysis in Table 3. The change of the relationship between supervisor's seniority and the Ph.D. student's productivity might result from the evolution of the mentorship style along the supervisor's career. Specifically, mid-career supervisors who intend to boost their productivity under the pressure of being promoted from associate to full professors might consider students as lab workforce involving them in several projects, even not directly linked to their thesis, and co-author publications with them (Mangematin and Robin, 2003; Shibayama, 2019). As a result, when looking at the overall number of students' publications (the analysis reported in Table 3), and at the number of publications co-authored with the supervisors (the analysis reported in Table C2, in Appendix C), we observe an inverted U-shaped relationship between the supervisor's seniority and the student's productivity. On the contrary, this relationship turns in a U-shaped form when we focus on students' publications similar to the thesis content (Table D2, in Appendix D). Indeed, students considered as lab workforce by mid-career supervisors might lower the number of articles related to their thesis subject in favor of articles related to the supervisors' projects.

1.8 Conclusion

Students, directors of Ph.D. programs, and policymakers urge to identify the environmental characteristics correlated to Ph.D. students’ productivity. From the students' perspective, showing a high-quality publication record and having a well-established scientific network is essential to be competitive in the job market after graduation. At the same time, directors of Ph.D. programs and policymakers need to optimize the use of resources and guarantee effective training programs.

In this paper, we study how social environment characteristics influence the Ph.D. students’ productivity during their training period using a dataset that covers the entire population of 77,143 Ph.D. students who graduated from French universities in STEM disciplines between 2000 and 2014.

We consider the supervisor and peers' biographic and academic characteristics as relevant social environment characteristics. Then, we measure the student’s productivity by counting the number of articles published during the training period (publication quantity), calculating the average number of citations received by the published articles (publication quality), and counting the number of distinct co-authors during the training period (scientific network size).

Not surprisingly, we find that students in productive environments are more productive, according to almost all the productivity measures considered. Having a female supervisor is associated with higher student productivity in engineering, the most male-centered discipline in our sample. Surprisingly, mentorship experience is associated with lower Ph.D. student’s productivity, while having a mid-career supervisor is associated with higher student productivity. Having a supervisor with a French or European research grant is associated with a higher number of citations received by the student. Sharing the training experience with large groups of peers penalizes student’s productivity, most likely due to a decline in the quality of the mentorship activity in large groups. On the contrary, having freshman peers, peers who publish high-quality articles, and at least one female peer is positively associated with student’s productivity.

Some of our results align with a recent survey conducted in France in 2021 involving more than eleven thousand Ph.D. students from all fields (Pommier et al., 2022). The survey aimed to explore the Ph.D. students’ perception of French Ph.D. programs. Results show that most respondents favor small-size teamwork with 2-3 students per supervisor. Consistently, half of the students declaring a lack of thesis progress is mentored by supervisors having more than four Ph.D. students simultaneously. Furthermore, the report shows that students evaluate the supervisor's role as fundamental in supporting the thesis progress and ensuring the financial conditions to carry out the research work.

A caveat applies to our analysis, as to a large part of the existing literature on Ph.D. students’ productivity. Our econometric approach does not strictly allow a causal interpretation of the relationships between dependent and independent variables in our regression exercises. Nonetheless, we believe that limited biases affect our estimates for three reasons. First, we reduce the omitted variable problem by including proxies for all the factors that the extant literature considers relevant in affecting students' productivity, such as supervisor, peers, department, and student's time-variant and time-invariant characteristics. Second, theory suggests that information asymmetry in student selection makes it unlikely to observe a correlation between students' unobserved intrinsic ability and supervisors' quality, which might generate a potential endogeneity issue (Mangematin, 2000). In line with the theory, the empirical literature shows a weak correlation between proxies for the student's ability and Ph.D. productivity (Belavy et al., 2020). Third, to further investigate the potential endogeneity issue, we included in our model specification a proxy for students' ability using data on the participation of the students in selective contests during high school. Including this variable does not affect our main results, confirming the low likelihood of biased estimates in our regression exercises.

Our results speak to Ph.D. students, directors of Ph.D. programs, and policymakers. On the one hand, our paper provides hints to the students who want to leverage the environmental factors to boost their productivity. On the other hand, our results provide directors of Ph.D. programs and policymakers with a framework to understand the determinants of effective training programs and find levers for designing policies that maximize students' productivity. Along these lines, our work can be exploited to design better Ph.D. programs. Using our regression estimates, we can simulate how the students' productivity varies according to environment characteristics' changes. For example, by increasing the supervisor's publications by one standard deviation, decreasing the number of peers by one student, and reducing the average experience of the supervisors by one standard deviation, we obtain that the student's predicted productivity increases by one publication, one citation, and four additional co-authors. According to these predictions and causally interpreting our regression results, we may suggest three policy interventions that can be applied in the short run to increase the effectiveness of the French Ph.D. training system. First, professors' requirements to access students' supervision might be revised. In France, professors who supervise Ph.D. students must obtain a habilitation, *Habilitation a Diriger des Recherches*. The habilitation is awarded mainly by looking at the professor's scientific achievements. Raising the threshold for obtaining the habilitation would ensure supervisors with a higher number of publications and, according to our results, more productive students[29]. Second, a rule limiting the number of supervised students might be introduced, reducing the average number of peers. Finally, scientists who have never mentored

[29] We assume that raising the threshold for the habitation does not create an undersupply of supervisors.

Ph.D. students and fulfilling the requirements to supervise should be incentivized to start the mentorship activity, reducing the overall average experience of the supervisors. Combining these three policy interventions would enhance the effectiveness of the current Ph.D. training programs.

1.9 Appendix of Chapter 1

APPENDIX A

This appendix reports the correlation matrix of the regressors included in Table 3. We find the highest correlation values between the variables *Supervisor's publications* and *Supervisor's co-authors* (0.78) and between *Peers' publications* and *Peers' co-authors* (0.89). In an alternative specification of the model estimated in Table 3, we excluded *Supervisor's* and *Peers' co-authors* from the model. Moreover, based on a Variance Inflation Factor (VIF) multicollinearity test, we excluded the variable *Citation-weighted publications per affiliate* that gives the highest VIF value (6.34). The model estimates excluding these three variables are consistent with those of Table 3 (Estimates without variables showing high correlation are available upon request).

Table A1. Variable correlation matrix (N=77,143)

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]
[1] Female supervisor	1.00																					
[2] Supervisor’s seniority	-0.01	1.00																				
[3] Mentorship experience	-0.09	0.12	1.00																			
[4] Supervisor’s publications	-0.11	0.32	0.19	1.00																		
[5] Supervisor avg. citations	0.03	0.24	-0.05	0.15	1.00																	
[6] Supervisor’s co-authors	-0.05	0.33	0.07	0.78	0.24	1.00																
[7] ANR grant	0.02	0.17	-0.02	0.09	0.15	0.11	1.00															
[8] EU grant	-0.03	0.03	0.01	0.11	0.06	0.10	0.02	1.00														
[9] With peers	-0.06	0.09	0.17	0.11	0.02	0.04	0.06	0.03	1.00													
[10] N. peers	-0.09	0.04	0.50	0.17	-0.03	0.03	0.03	0.02	0.41	1.00												
[11] At least one female peer	0.00	0.09	0.19	0.14	0.05	0.09	0.06	0.03	0.52	0.44	1.00											
[12] Average peers’ seniority	-0.08	0.13	0.26	0.14	0.01	0.05	0.05	0.04	0.77	0.56	0.52	1.00										
[13] Peers’ publications	-0.04	0.10	0.20	0.27	0.05	0.17	0.04	0.03	0.23	0.45	0.22	0.32	1.00									
[14] Peers’ average citations	-0.03	0.10	0.12	0.24	0.14	0.20	0.08	0.04	0.17	0.32	0.17	0.24	0.76	1.00								
[15] Peers’ co-authors	-0.03	0.11	0.14	0.24	0.08	0.22	0.05	0.03	0.21	0.37	0.20	0.28	0.89	0.76	1.00							
[16] French Top-20	0.06	0.02	0.03	0.04	0.08	0.05	0.04	0.04	0.00	0.04	0.05	0.00	0.04	0.04	0.03	1.00						
[17] Citation-weighted publications per affiliate	0.09	0.40	-0.09	0.14	0.26	0.25	0.20	0.01	-0.04	-0.12	0.02	-0.05	0.04	0.08	0.07	0.06	1.00					
[18] IDEX	0.05	0.27	0.01	0.02	0.14	0.09	0.18	-0.02	0.01	0.01	0.03	0.03	0.04	0.06	0.05	0.14	0.44	1.00				
[19] Department size [100 aff.]	0.14	0.17	-0.06	0.18	0.21	0.25	0.10	0.04	-0.07	-0.11	0.05	-0.08	0.05	0.09	0.08	0.33	0.50	0.24	1.00			
[20] N. Ph.D. stud. in program	0.07	0.20	-0.02	0.11	0.19	0.15	0.11	0.05	0.02	0.04	0.05	0.03	0.08	0.10	0.08	0.33	0.38	0.29	0.47	1.00		
[21] Female student	0.09	0.04	-0.03	0.05	0.06	0.08	0.02	0.00	-0.04	-0.06	0.05	-0.04	-0.01	0.00	0.01	0.06	0.11	0.03	0.18	0.05	1.00	
[22] Co-supervision	0.00	0.13	0.00	-0.02	0.00	-0.02	0.04	-0.02	0.00	0.00	-0.01	0.03	-0.01	-0.01	-0.01	-0.12	0.08	0.06	-0.14	-0.04	-0.01	1.00

APPENDIX B

Table B1. Descriptive Statistics of the explanatory variables, by field.

	Engineering				Mathematics				Medicine-biology-chemistry				Physics			
77,143 Ph.D. students	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Supervisor characteristics																
Female supervisor	0.14	0.35	0.00	1.00	0.17	0.37	0.00	1.00	0.28	0.45	0.00	1.00	0.17	0.38	0.00	1.00
Supervisor’s seniority	11.11	5.07	0.00	21.00	9.89	5.46	0.00	21.00	12.20	4.99	0.00	21.00	11.47	5.53	0.00	21.00
Mentorship experience	4.41	7.44	0.00	114.00	3.97	7.48	0.00	114.00	2.37	5.49	0.00	184.00	2.56	4.71	0.00	108.0
Supervisor’s publications	11.01	11.93	0.00	98.00	6.92	9.46	0.00	93.00	16.86	15.69	0.00	100.00	13.91	14.07	0.00	100.0
Supervisor’s average citations	1.76	2.27	0.00	87.17	1.54	3.58	0.00	127.87	2.95	3.08	0.00	113.09	2.28	2.88	0.00	98.22
Supervisor’s co-authors	22.72	34.31	0.00	498.00	13.08	29.38	0.00	468.00	50.82	56.15	0.00	499.0	40.36	54.95	0.00	498.00
ANR grant	0.05	0.21	0.00	1.00	0.04	0.20	0.00	1.00	0.08	0.28	0.00	1.00	0.06	0.23	0.00	1.00
EU grant	0.02	0.14	0.00	1.00	0.01	0.12	0.00	1.00	0.03	0.16	0.00	1.00	0.03	0.18	0.00	1.00
Peer characteristics																
With peers	0.89	0.31	0.00	1.00	0.84	0.37	0.00	1.00	0.76	0.43	0.00	1.00	0.77	0.42	0.00	1.00
N. peers	2.54	2.48	0.00	28.25	2.27	2.73	0.00	28.25	1.33	1.68	0.00	28.25	1.49	1.80	0.00	30.00
At least one female peer	0.53	0.50	0.00	1.00	0.48	0.50	0.00	1.00	0.55	0.50	0.00	1.00	0.45	0.50	0.00	1.00
Average peers’ seniority	1.91	0.91	0.00	3.48	1.76	1.00	0.00	3.43	1.46	1.07	0.00	3.44	1.51	1.06	0.00	3.56
Peers’ publications	0.88	1.94	0.00	27.25	0.68	1.69	0.00	29.75	0.85	1.78	0.00	41.00	0.70	1.55	0.00	25.75
Peers’ average citations	2.57	8.41	0.00	353.15	1.88	7.38	0.00	187.40	3.15	8.56	0.00	266.58	2.47	6.99	0.00	150.54
Peers’ co-authors	4.21	10.87	0.00	190.75	3.18	9.60	0.00	176.25	4.77	10.54	0.00	187.25	3.66	9.29	0.00	150.00
Other controls																
French Top-20	0.24	0.43	0.00	1.00	0.53	0.50	0.00	1.00	0.49	0.50	0.00	1.00	0.19	0.39	0.00	1.00
Citation-weighted publications per affiliate	3.96	1.61	0.38	10.72	3.71	1.55	0.81	10.61	8.54	3.41	0.93	17.58	11.43	5.40	1.35	35.05
IDEX	0.15	0.36	0.00	1.00	0.19	0.39	0.00	1.00	0.19	0.39	0.00	1.00	0.19	0.40	0.00	1.00
Department size [100 affiliates]	9.54	6.12	0.04	27.99	6.57	4.55	0.10	21.54	49.33	32.74	0.18	114.46	20.87	18.64	0.15	64.30
N. of Ph.D. students in the program	753.04	680.93	5.00	2973.0	1000.73	795.82	1.00	2973.0	1138.96	803.44	1.00	2973.0	1173.13	840.62	1.00	2973.0
Female student	0.25	0.43	0.00	1.00	0.27	0.44	0.00	1.00	0.53	0.50	0.00	1.00	0.33	0.47	0.00	1.00
Co-supervision	0.39	0.49	0.00	1.00	0.31	0.46	0.00	1.00	0.26	0.44	0.00	1.00	0.35	0.48	0.00	1.00
Entry year	2005.20	4.13	1997.0	2011.0	2005.47	4.08	1997.0	2011.0	2004.93	4.21	1997.0	2011.0	2005.23	4.30	1997.00	2011.0
Observations	16,519				11,450				35,038				14,136			

APPENDIX C

This appendix reports a robustness check in which we only select the publications of the Ph.D. student co-authored with her supervisor to calculate our dependent variables. Using this selection criterion, we find that 59.79% of students have at least one paper co-authored with the supervisor during the training period.

Table C1 shows the descriptive statistics of the newly calculated dependent variables, while Table C2 shows the regression results.

Table C1. Descriptive statistics of the students' productivity outcomes. Publication attribution is based on the co-authorship with the supervisor.

Dependent variables 77,143 Ph.D. students	Mean	Sd	Min	Max
Publications	1.76	2.33	0.00	20.00
Average citations	1.97	3.59	0.00	170.42
Co-authors	6.92	12.27	0.00	195.00

Table C2. Regression results. Publication attribution is based on the co-authorship with the supervisor. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Supervisor characteristics			
Female supervisor	0.028	0.069**	0.37***
	(0.019)	(0.030)	(0.10)
Supervisor’s seniority	0.11***	0.067***	0.33***
	(0.0055)	(0.0086)	(0.029)
Supervisor’s seniority^2	-0.0048***	-0.0032***	-0.015***
	(0.00026)	(0.00041)	(0.0014)
Mentorship experience	-0.019***	-0.0081***	-0.041***
	(0.0015)	(0.0023)	(0.0076)
Supervisor’s publications	0.027***	0.0078***	-0.083***
	(0.00094)	(0.0015)	(0.0049)
Supervisor’s average citations	0.038***	0.21***	0.23***
	(0.0028)	(0.0043)	(0.014)
Supervisor’s co-authors	0.00073***	0.0019***	0.074***
	(0.00026)	(0.00041)	(0.0013)
ANR grant	0.16***	0.58***	0.71***
	(0.033)	(0.051)	(0.17)
EU grant	-0.15***	0.27***	-0.94***
	(0.050)	(0.078)	(0.26)
Peer characteristics			
With peers	0.20***	0.23***	0.53***
	(0.031)	(0.049)	(0.16)
N. peers	-0.077***	-0.046***	-0.28***
	(0.0054)	(0.0085)	(0.028)
At least one female peer	-0.026	0.071**	0.16
	(0.019)	(0.030)	(0.100)
Average peers’ seniority	-0.16***	-0.14***	-0.64***
	(0.013)	(0.021)	(0.068)
Peers’ publications	0.074***	-0.17***	-0.67***
	(0.011)	(0.017)	(0.055)
Peers’ average citations	0.011***	0.059***	0.067***
	(0.0015)	(0.0024)	(0.0080)
Peers’ co-authors	0.0011	0.0039	0.17***
	(0.0017)	(0.0027)	(0.0090)

Other controls			
French Top-20	-0.063***	0.044	-0.40***
	(0.018)	(0.028)	(0.093)
Citation-weighted publications per affiliate	0.013***	0.026***	0.13***
	(0.0043)	(0.0068)	(0.023)
IDEX	-0.021	0.0088	0.035
	(0.027)	(0.043)	(0.14)
Department size [100 affiliates]	-0.00042	0.0016**	0.0057**
	(0.00043)	(0.00068)	(0.0023)
N. of Ph.D. students in the program	0.000047***	0.00020***	0.00028***
	(0.000012)	(0.000019)	(0.000062)
Female student	-0.36***	-0.18***	-1.05***
	(0.016)	(0.026)	(0.084)
Co-supervision	-0.094***	-0.077***	-0.60***
	(0.018)	(0.028)	(0.091)
Engineering	0.35***	0.45***	0.92***
	(0.027)	(0.042)	(0.14)
Physics	0.72***	0.61***	1.78***
	(0.042)	(0.066)	(0.22)
Medicine-biology-chemistry	1.41***	1.44***	5.24***
	(0.033)	(0.051)	(0.17)
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	0.25***	-0.25***	1.14***
	(0.056)	(0.088)	(0.29)
Observations	77,143	77,143	77,143
R-squared	0.172	0.135	0.193

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. Standard errors are reported in parentheses.

APPENDIX D

This appendix reports a robustness check in which we only select the Ph.D. student’s publications showing high similarity with the abstract of the thesis manuscript. We expect that a large part of students' publications during the training period derives from the thesis research work.

To measure the similarity between the student’s thesis and her publications, we rely on a text analysis algorithm (Mikolov et al., 2013). Specifically, we proceed in two steps. First, we use the Word2Vec algorithm to attribute to each word its vectorial representation according to the word’s semantic meaning. To do so, we generate a vocabulary of 411,525 words retrieved from all the distinct words appearing in the abstracts of 1,284,753 STEM publications in English authored by French researchers in 1990-2018. Then, we use the co-occurrence of words in the articles’ abstract to train our algorithm and provide for each word a 100-dimension vectorial representation (Rong, 2014). Each dimension in the vectorial space represents a latent dimension of the word's semantic meaning. Once we generate a vocabulary that allows us to translate words into vectors, we attribute to each word appearing within theses and student publications’ abstracts its vectorial representation. Therefore, after this operation, theses and students’ publications are represented by a series of vectors corresponding to words, each of which is a point in the 100-dimension vectorial space. In order to obtain the vectorial representation of the entire text documents, we calculate the centroid of al the vectors representing each document. When all the documents are represented by a unique vector, we calculate the cosine similarity between the vectors representing the student thesis and the vectors representing the student’s publications. Cosine similarity values range from -1 (highly dissimilar documents) to +1 (highly similar documents). We consider a thesis similar to a publication if the cosine similarity value exceeds the threshold of 0.8. Once calculated the similarity between documents, we attribute to students only papers similar to her Ph.D. thesis. We end up with 44.27% of students having at least one paper attributed.

Table D1 shows the descriptive statistics of the newly calculated dependent variables, while Table D2 shows the regression results.

Table D1. Descriptive statistics of the students' productivity outcomes. Publication attribution is based on similarity between student’s thesis and publications.

Dependent variables 77,143 Ph.D. students	Mean	Sd	Min	Max
Publications	1.38	2.30	0.00	20.00
Average citations	1.41	3.09	0.00	120.24
Co-authors	5.37	11.82	0.00	200.00

Table D2. Regression results. Publication attribution is based on similarity between student's thesis and publications. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Supervisor characteristics			
Female supervisor	0.035*	0.087***	0.33***
	(0.019)	(0.026)	(0.098)
Supervisor's seniority	-0.015***	-0.033***	-0.085***
	(0.0055)	(0.0075)	(0.028)
Supervisor's seniority^2	0.00075***	0.0012***	0.0042***
	(0.00026)	(0.00036)	(0.0013)
Mentorship experience	-0.011***	-0.0028	-0.019**
	(0.0015)	(0.0020)	(0.0075)
Supervisor's publications	0.016***	0.0015	-0.078***
	(0.00094)	(0.0013)	(0.0048)
Supervisor's average citations	0.025***	0.14***	0.17***
	(0.0028)	(0.0038)	(0.014)
Supervisor's co-authors	0.0013***	0.0016***	0.058***
	(0.00026)	(0.00036)	(0.0013)
ANR grant	0.19***	0.69***	0.99***
	(0.033)	(0.045)	(0.17)
EU grant	-0.12**	0.15**	-0.85***
	(0.050)	(0.068)	(0.25)
Peer characteristics			
With peers	0.16***	0.14***	0.53***
	(0.031)	(0.043)	(0.16)
N. peers	-0.072***	-0.035***	-0.24***
	(0.0054)	(0.0074)	(0.027)
At least one female peer	-0.012	0.056**	0.16
	(0.019)	(0.026)	(0.098)
Average peers' seniority	-0.092***	-0.063***	-0.49***
	(0.013)	(0.018)	(0.066)
Peers' publications	0.086***	-0.089***	-0.30***
	(0.011)	(0.015)	(0.054)
Peers' average citations	-0.00032	0.028***	0.0048
	(0.0015)	(0.0021)	(0.0078)
Peers' co-authors	0.0013	0.0038	0.12***
	(0.0017)	(0.0024)	(0.0089)
Other controls			
French Top-20	-0.24***	-0.17***	-1.00***
	(0.018)	(0.024)	(0.091)
Citation-weighted publications per affiliate	0.070***	0.073***	0.36***
	(0.0044)	(0.0060)	(0.022)
IDEX	0.050*	0.19***	0.59***
	(0.027)	(0.037)	(0.14)
Department size [100 affiliates]	-0.0032***	-0.0038***	-0.012***
	(0.00043)	(0.00059)	(0.0022)
N. of Ph.D. students in the program	-0.00017***	-0.000097***	-0.00054***
	(0.000012)	(0.000016)	(0.000061)
Female student	-0.32***	-0.15***	-0.96***
	(0.016)	(0.022)	(0.083)
Co-supervision	0.076***	0.044*	0.0055
	(0.018)	(0.024)	(0.089)
Engineering	0.19***	0.33***	0.79***
	(0.027)	(0.037)	(0.14)
Physics	0.14***	0.034	0.27
	(0.042)	(0.058)	(0.21)
Medicine-biology-chemistry	0.69***	0.85***	3.42***
	(0.033)	(0.045)	(0.17)
Mathematics	Ref.	Ref.	Ref.

Entry year dummies	Yes	Yes	Yes
Constant	1.19***	0.72***	3.66***
	(0.056)	(0.077)	(0.29)
Observations	77,143	77,143	77,143
R-squared	0.146	0.114	0.165

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. Standard errors are reported in parentheses.

APPENDIX E

This appendix reports a regression exercise where we include a proxy for the student's intrinsic ability among the control variables. Specifically, we identify in our study sample the students who have participated in three well-known contests during the high school period: the International Mathematical Olympiad (IMO), *Les Olympiades Nationales de Mathématiques* (the national French Mathematical Olympiad), and *le Kangourou des mathématiques* (a French national mathematical contest)[30]. These contests are organized both at the national and international level, and students who show particular abilities during their high school studies are selected to participate. We argue that this variable is a good proxy for students' intrinsic ability, interest, and motivation in schooling and education.

We found 138 Ph.D. students who participated in at least one of the three contests and were mentioned in the contests' final ranking (with or without winning a medal). In our econometric exercise, we identify those students with the dummy variable *Math Olympiad* that equals one if the student participated in at least one of the three contests, zero otherwise. As expected, we find that a large share of students ends up doing a Ph.D. in Mathematics (53%); nonetheless, a non-negligible share did a Ph.D. in engineering (19%), Physics (12%), and Medicine-biology-chemistry (16%).

Table E1 reports the regression exercise results, including the *Math Olympiad* dummy variable among the controls. The results concerning the supervisor's and peers' characteristics are in line with those presented in Table 3 in our main analysis, and the dummy *Math Olympiad* is never significant in all the three econometric models considered.

We conclude that including a proxy for the student's ability does not change the impact of the environmental characteristics on the student's scientific productivity. These results are consistent with previous literature findings (Aristizábal, 2021; Belavy et al., 2020; Mangematin, 2000).

[30] Data for the International Mathematical Olympiad (IMO) are available from 1981 to 2009, for *Les Olympiades Nationales de Mathématiques* from 2001 to 2007, and for *le Kangourou des mathématiques* from 2005 to 2007.

Table E1. Regression results. Including a proxy for the student's ability. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Student's ability			
Math Olympiad	0.19	-0.0094	-0.87
	(0.24)	(0.28)	(1.19)
Supervisor characteristics			
Female supervisor	-0.0049	0.074**	0.31**
	(0.025)	(0.030)	(0.13)
Supervisor's seniority	0.037***	0.0071	0.11***
	(0.0072)	(0.0085)	(0.036)
Supervisor's seniority^2	-0.0019***	-0.00096**	-0.0067***
	(0.00034)	(0.00040)	(0.0017)
Mentorship experience	-0.018***	-0.0072***	-0.037***
	(0.0019)	(0.0023)	(0.0097)
Supervisor's publications	0.027***	0.0070***	-0.10***
	(0.0012)	(0.0015)	(0.0062)
Supervisor's average citations	0.031***	0.20***	0.21***
	(0.0036)	(0.0043)	(0.018)
Supervisor's co-authors	0.0028***	0.0014***	0.091***
	(0.00034)	(0.00040)	(0.0017)
ANR grant	0.0050	0.54***	0.22
	(0.043)	(0.050)	(0.21)
EU grant	-0.19***	0.33***	-1.28***
	(0.065)	(0.077)	(0.33)
Peer characteristics			
With peers	0.13***	0.24***	0.25
	(0.041)	(0.048)	(0.21)
N. peers	-0.12***	-0.042***	-0.39***
	(0.0071)	(0.0083)	(0.036)
At least one female peer	-0.028	0.073**	0.21*
	(0.025)	(0.030)	(0.13)
Average peers' seniority	-0.14***	-0.13***	-0.63***
	(0.017)	(0.020)	(0.086)
Peers' publications	0.13***	-0.15***	-0.64***
	(0.014)	(0.016)	(0.070)
Peers' average citations	0.0065***	0.056***	0.049***
	(0.0020)	(0.0024)	(0.010)
Peers' co-authors	0.0029	0.0017	0.21***
	(0.0023)	(0.0027)	(0.011)
Other controls			
French Top-20	-0.0084	0.068**	-0.36***
	(0.023)	(0.028)	(0.12)
Citation-weighted publications per affiliate	0.012**	0.026***	0.14***
	(0.0057)	(0.0067)	(0.029)
IDEX	-0.056	0.031	-0.031
	(0.036)	(0.042)	(0.18)
Department size [100 affiliates]	0.00081	0.0014**	0.013***
	(0.00057)	(0.00067)	(0.0029)
N. of Ph.D. students in the program	0.000092***	0.00023***	0.00038***
	(0.000016)	(0.000019)	(0.000079)
Female student	-0.64***	-0.19***	-1.84***
	(0.021)	(0.025)	(0.11)
Co-supervision	-0.065***	-0.042	-0.66***
	(0.023)	(0.027)	(0.12)
Engineering	0.18***	0.40***	0.99***
	(0.035)	(0.041)	(0.18)
Physics	0.77***	0.57***	2.45***
	(0.055)	(0.065)	(0.28)
Medicine-biology-chemistry	1.54***	1.39***	6.44***

	(0.043)	(0.050)	(0.22)
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	1.23***	0.23***	3.84***
	(0.074)	(0.087)	(0.37)
Observations	77,143	77,143	77,143
R-squared	0.140	0.128	0.174

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. Standard errors are reported in parentheses.

APPENDIX F

This appendix searches for non-linear associations between student outcomes, number of peers, and supervisor's mentorship experience.

Table F1 investigates the possible nonlinear association between the supervisor's mentorship experience and the student's productivity outcomes. We calculate eight dummy variables, each of which takes value one if the number of Ph.D. students successfully supervised before the focal student enrollment equals 1, 2, 3, 4, 5, 6, 7, or is larger than 7. Specifically, The variable *Mentorship experience=1*, takes value one if the supervisor has only one Ph.D. student who graduated before the enrollment of the focal student. The variable equals one for 15.1% of supervisors. Similarly, we calculate *Mentorship experience=2* (9.7%), *Mentorship experience=3* (7.0%), *Mentorship experience=4* (5.1%), *Mentorship experience=5* (3.9%), *Mentorship experience=6* (2.9%), *Mentorship experience=7* (2.4%), and *Mentorship experience>7* (11.7%). The reference case, represented by the vertical line centered in zero, is when the supervisor has *No mentorship experience* (42.1%). We find that results in Table F1 confirm a negative association between the supervisor's mentorship experience and the three student's outcomes, as shown in Table 3. Figure 1 in the main text depicts the association between mentorship experience and Ph.D. students' productivity.

Table F2 investigates the possible nonlinear association between the student's number of peers and the student's productivity outcomes. Based on the values of the variables *N. peers*, we calculated six dummy variables. The first variable, *0 < N. peers <=1*, takes value one for all those students having between 0 (excluded) and 1 (included) peers during the Ph.D. period. The variable equals one for 29.9% of the students. Similarly, we calculate *1 < N. peers <=2* (20.7% of the students), *2 < N. peers <=3* (12.2%), *3 < N. peers <=4* (6.9%), *4 < N. peers <=5* (3.8%), and *N. peers >5* (6.4%). The reference case is when the focal student has *No peers* (20.0%). Table F2 shows a similar pattern as the one observed in the regressions in Table 3 in the main text. We find a positive association between *0 < N. peers <=1* on the citations received by the student's articles. An increase in the number of peers is associated with a sharp decrease of the student's publications leading to -0.83 articles and -3.21 co-authors when the peer number exceeds 5 peers. Interestingly, a large number of peers is not associated with a significant decrease in citations.

We interpret the result on publication productivity as a loss of supervisor's attention to the student's work. In the case of many peers, the supervisor shares her limited time with many students reducing her support to each of them. A similar interpretation can apply to the citations received by the student's work. If the supervisor has up to 3 students at a time (the focal student + 2 peers) the quality of the student's work probably benefits from the supervisor's advice. Concerning the negative association between the number of peers and the number of co-authors, one possible explanation is that the student having many peers within the team has less incentive to look for other collaborators outside the team, reducing the probability of finding new co-authors or joining other research teams. Figure 2 in the main text provides a visual representation of the association between the peer group size (*N. peers*) and the Ph.D. student's productivity.

Table F1. Regression with mentoring experience dummy variables.

VARIABLES	(1) Publications	(2) Average citations	(3) Co-authors
Supervisor Female	-0.010	0.071**	0.29**
	(0.025)	(0.030)	(0.13)
Supervisor's Seniority	0.033***	0.0046	0.094***
	(0.0072)	(0.0085)	(0.036)
Supervisor's Seniority^2	-0.0014***	-0.00073*	-0.0050***
	(0.00034)	(0.00040)	(0.0017)
Mentorship experience = 0	**Ref.**	**Ref.**	**Ref.**
Mentorship experience = 1	**-0.074****	**0.012**	**-0.0049**
	(0.031)	**(0.036)**	**(0.15)**
Mentorship experience = 2	**-0.23*****	**-0.100****	**-0.48*****
	(0.037)	**(0.043)**	**(0.18)**
Mentorship experience = 3	**-0.26*****	**-0.070**	**-0.73*****
	(0.042)	**(0.050)**	**(0.21)**
Mentorship experience = 4	**-0.26*****	**-0.062**	**-0.65*****
	(0.048)	**(0.057)**	**(0.24)**
Mentorship experience = 5	**-0.38*****	**-0.18*****	**-1.14*****
	(0.054)	**(0.064)**	**(0.27)**
Mentorship experience = 6	**-0.32*****	**-0.14***	**-1.16*****
	(0.062)	**(0.073)**	**(0.31)**
Mentorship experience = 7	**-0.52*****	**-0.19****	**-1.76*****
	(0.068)	**(0.080)**	**(0.34)**
Mentorship experience > 7	**-0.55*****	**-0.25*****	**-1.44*****
	(0.039)	**(0.046)**	**(0.19)**
Supervisors' Publications	0.028***	0.0077***	-0.099***
	(0.0012)	(0.0015)	(0.0062)
Supervisors' Average citations	0.030***	0.20***	0.21***
	(0.0036)	(0.0043)	(0.018)
Supervisors' Co-authors	0.0026***	0.0013***	0.090***
	(0.00034)	(0.00040)	(0.0017)
ANR grant	0.0013	0.53***	0.20
	(0.043)	(0.050)	(0.21)
EU grant	-0.17***	0.34***	-1.23***
	(0.065)	(0.077)	(0.33)
With peers	0.11***	0.23***	0.19
	(0.041)	(0.048)	(0.21)
N. peers	-0.12***	-0.040***	-0.37***
	(0.0068)	(0.0080)	(0.034)
At least one female peer	-0.015	0.079***	0.25**
	(0.025)	(0.030)	(0.13)
Average peers' seniority	-0.10***	-0.12***	-0.53***
	(0.017)	(0.020)	(0.087)
Peers' publications	0.13***	-0.15***	-0.64***
	(0.014)	(0.016)	(0.070)
Peers' average citations	0.0067***	0.056***	0.049***
	(0.0020)	(0.0024)	(0.010)
Peers' co-authors	0.0034	0.0019	0.21***
	(0.0023)	(0.0027)	(0.011)
French Top-20	-0.0080	0.069**	-0.36***
	(0.023)	(0.028)	(0.12)
Citation-weighted publications per affiliate	0.011*	0.026***	0.14***
	(0.0057)	(0.0067)	(0.029)
IDEX	-0.039	0.038	0.023
	(0.036)	(0.042)	(0.18)
Department size [100 affiliates]	0.0011*	0.0015**	0.014***
	(0.00057)	(0.00067)	(0.0029)
N. of Ph.D. students in the program	0.000074***	0.00022***	0.00033***
	(0.000016)	(0.000019)	(0.000079)

Female student	-0.64***	-0.19***	-1.84***
	(0.021)	(0.025)	(0.11)
Co-supervision	-0.066***	-0.041	-0.66***
	(0.023)	(0.027)	(0.12)
Engineering	0.17***	0.39***	0.96***
	(0.035)	(0.041)	(0.18)
Physics	0.76***	0.56***	2.41***
	(0.055)	(0.065)	(0.28)
Medicine-biology-chemistry	1.50***	1.37***	6.32***
	(0.043)	(0.051)	(0.22)
Mathematics	Ref.	Ref.	Ref.
Constant	1.54***	-0.12	2.51***
	(0.068)	(0.080)	(0.34)
Observations	77,143	77,143	77,143
R-squared	0.142	0.128	0.175
Entry year dummies	Yes	Yes	Yes

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. Standard errors are reported in parentheses.

Table F2. Regression with dummy variables for the peer group size.

	(1)	(2)	(3)
VARIABLES	Publications	Average citations	Coauthors
Supervisor Female	-0.0073	0.074**	0.29**
	(0.025)	(0.030)	(0.13)
Supervisor's Seniority	0.036***	0.0068	0.11***
	(0.0072)	(0.0085)	(0.036)
Supervisor's Seniority^2	-0.0019***	-0.00094**	-0.0066***
	(0.00034)	(0.00040)	(0.0017)
Mentorship experience	-0.022***	-0.0086***	-0.052***
	(0.0018)	(0.0022)	(0.0093)
Supervisors' Publications	0.027***	0.0072***	-0.10***
	(0.0012)	(0.0015)	(0.0062)
Supervisors' Average citations	0.031***	0.20***	0.21***
	(0.0036)	(0.0043)	(0.018)
Supervisors' Co-authors	0.0027***	0.0014***	0.090***
	(0.00034)	(0.00040)	(0.0017)
ANR grant	0.0039	0.54***	0.23
	(0.043)	(0.050)	(0.21)
EU grant	-0.18***	0.33***	-1.22***
	(0.065)	(0.077)	(0.33)
No peers	**Ref.**	**Ref.**	**Ref.**
0 < N. peers <=1	**0.048**	**0.21*** **	**-0.12**
	(0.042)	**(0.049)**	**(0.21)**
1 < N. peers <=2	**-0.065**	**0.22*** **	**-0.79*** **
	(0.057)	**(0.067)**	**(0.28)**
2 < N. peers <=3	**-0.27*** **	**0.12**	**-1.65*** **
	(0.063)	**(0.074)**	**(0.32)**
3 < N. peers <=4	**-0.41*** **	**0.032**	**-2.13*** **
	(0.069)	**(0.081)**	**(0.35)**
4 < N. peers <=5	**-0.58*** **	**-0.060**	**-2.58*** **
	(0.078)	**(0.092)**	**(0.39)**
N. peers >5	**-0.83*** **	**-0.060**	**-3.21*** **
	(0.074)	**(0.088)**	**(0.37)**
At least one female peer	-0.0045	0.082***	0.37***
	(0.026)	(0.030)	(0.13)
Average peers' seniority	-0.12***	-0.13***	-0.45***
	(0.019)	(0.023)	(0.098)
Peers' publications	0.12***	-0.15***	-0.65***

	(0.014)	(0.016)	(0.069)
Peers' average citations	0.0070***	0.056***	0.051***
	(0.0020)	(0.0024)	(0.010)
Peers' co-authors	0.0035	0.0018	0.21***
	(0.0023)	(0.0027)	(0.011)
French Top-20	-0.014	0.067**	-0.38***
	(0.023)	(0.028)	(0.12)
Citation-weighted publications per affiliate	0.012**	0.026***	0.14***
	(0.0057)	(0.0067)	(0.029)
IDEX	-0.055	0.030	-0.032
	(0.036)	(0.042)	(0.18)
Department size [100 affiliates]	0.00091	0.0014**	0.013***
	(0.00057)	(0.00067)	(0.0029)
N. of Ph.D. students in the program	0.000084***	0.00022***	0.00036***
	(0.000016)	(0.000019)	(0.000079)
Female student	-0.64***	-0.19***	-1.85***
	(0.021)	(0.025)	(0.11)
Co-supervision	-0.062***	-0.040	-0.65***
	(0.023)	(0.027)	(0.12)
Engineering	0.18***	0.40***	1.04***
	(0.035)	(0.041)	(0.18)
Physics	0.77***	0.57***	2.42***
	(0.055)	(0.065)	(0.28)
Medicine-biology-chemistry	1.52***	1.38***	6.36***
	(0.043)	(0.051)	(0.22)
Mathematics	Ref.	Ref.	Ref.
Constant	1.44***	-0.16**	2.30***
	(0.067)	(0.079)	(0.34)
Observations	77,143	77,143	77,143
R-squared	0.141	0.128	0.175
Entry year dummies	Yes	Yes	Yes

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. Standard errors are reported in parentheses.

CHAPTER 2

The impact of initiative of excellence funding on French researchers' outcomes

2.1 Introduction

Despite the increasing relevance of different funding sources[31], government funding remains the primary source of funds for universities in Europe[32] (OECD, 2019). Over the past fifteen years, governments have been increasingly relying on direct government funds to allocate financial resources to universities at the expense of the more traditional block funding (or general university funding). While block funds are distributed on a formula bases, direct government funds are based on a contract between universities and the government. This change, described as a turn in the rationale for science funding towards a contractual-oriented approach (Geuna, 2001), implies quasi-market incentives for universities that are asked to compete with each other for a stipulated amount of money to be used for specific objectives set by the government. In doing so, governments aim to influence the universities' research agenda according to the modern role they attribute to them. Universities are seen as actors in the country's economic development and international competitiveness and are responsible for addressing the new societal challenges, also through the transfer of applied knowledge to local and regional industries (Weber and Rohracher, 2012; Geuna and Rossi, 2015; Mazzucato, 2018; Valero and Van Reenen, 2019).

The preference for direct government funds has raised many concerns (Geuna, 2001; Stephan, 2012). Research constraints imposed on universities' agendas, competition for funding allocation based on performance, and "stop and go" funding depending on contracts are likely to harm universities' research potential, discouraging long-term research projects, and forcing universities to focus on short-term quantifiable results. Moreover, direct government funds introduce the major issue of measuring universities' performance. The multiplicity of activities in which universities are engaged, grouped in education, research, and technology transfer, combined with the limitation of available public data, makes it difficult to assess the government funding impact on universities' outcomes (Sarrico et al., 2010; Geuna and Rossi, 2015). Universities themselves are increasingly asked to be more accountable for what concerns funding (Geuna and Martin, 2003). From a social perspective, it is fundamental to assess the return of public spending to universities using a comprehensive approach.

The extant literature trying to quantify the impact of research funding is scant and does not provide convergent results. Moreover, it mainly focuses on grants awarded to individual researchers, neglecting funding programs addressed to universities (Arora and Gambardella, 2005; Jacob and Lefgren, 2011; Lane and Bertuzzi, 2011; Heyard and Hottenrott, 2021). The extant few studies on

---

[31] Funds from industry, private non-profit foundations, the European Commission, and university self-funding have raised in the share of universities' income over the past fifteen years (Stephan, 2012; OECD, 2019).

[32] In the EU23 countries, in 2016, 76.60% of universities' total expenditures are sourced from the public sector, on average. For the OECD countries the value decreases to 68.54% (OECD, 2019).

universities differ in the outcomes analyzed due to the difficulty in collecting data. Furthermore, they focus on one specific outcome at a time, overlooking the multifaced aspects of universities' activities, and do not provide a clear view of the extent to which government funding affects universities (Payne and Siow, 2003; Gawellek and Sunder, 2016; Wahls, 2018).

This paper aims to assess the impact of a direct government fund on a broad range of university researchers' activities. We focus on France, studying the case of IDEX (*Initiative D'EXcellence*) funding. IDEX is a competitive funding program launched by the French government in 2011 as a part of the *Plan d'Investissements pour l'Avenir* (PIA) national fiscal stimulus. Its main goal is to provide a selected group of French universities with resources to reach excellence at the global level and compete with leading universities worldwide. To do so, it favors the concentration of the French higher education and research institutions by providing funds to renovate buildings and facilities, and stimulates a broad range of researchers' activities that involve research, technology transfer, mentoring Ph.D. students, and fundraising. We analyze the impact of IDEX funding on researchers from seventeen applicant universities, eight awarded universities, and nine non-applicant universities. We consider eleven aspects of French researchers' activities that can be traced back to the goals of IDEX: quantity, quality, and interdisciplinarity of the publication productivity, interdisciplinarity in the collaborative behavior, scientific collaborations within the laboratory, scientific collaborations within the university, national scientific collaborations, international scientific collaborations, patenting, mentoring Ph.D. students, and fundraising.

We contribute to the funding literature by analyzing for the first time the impact of a direct government fund on a broad set of university researchers' outcomes that consider research, education, and technology transfer activities. Furthermore, we investigate the effect of applying for IDEX on researchers' activities, regardless of the application result. Gawellek and Sunder (2016) evidenced how German universities lost considerably in terms of efficiency during the application phase for government funds of excellence, due to the effort in preparing the proposal. Moreover, recent studies at the researcher level showed that limiting the scope of the study to awarded individuals introduces (selection) biases in the results and neglect a fundamental step in the funding process, namely the decision to apply for funding (Ayoubi et al. 2019; Davies et al., 2022). Finally, we also explore the possibility of funding spillovers[33], investigating the indirect impact of IDEX funding on researchers affiliated with universities that did not apply for IDEX but who collaborate with researchers affiliated

[33] We assume that spillovers in co-authorship networks occur due to knowledge exchanges between researchers who are co-authors of a scientific work. They are driven both by spontaneous social interactions and contractual agreements between researchers (Breschi and Lissoni, 2009).

with universities awarded IDEX. Highly connected scientific networks, showing ‘small world’ properties (Ebadi and Schiffauerova, 2015b), are likely to generate indirect effects of funding.

We estimate the effect of IDEX by analyzing a panel of 32,947 researchers in STEM disciplines observed between 2006 and 2015. We find that both being affiliated with a university that applies for IDEX and being affiliated with a university that is awarded IDEX affect researchers’ outcomes, but the effect is limited to the researchers’ network. Specifically, researchers increase their collaborations within their university (both within and outside their lab) and gain new national (outside their university) and international co-authors. The gain of national (+0.35 co-authors per year) and international co-authors (+1.06 co-authors per year) is solely due to the IDEX award. Moreover, we find positive network spillovers of IDEX for researchers in universities that did not apply for funding. If researchers in non-applicant universities are connected with awarded researchers, they increase their number of co-authors within the laboratory, at the national level (outside their university), and at the international level. When we break down our analysis by field, we find heterogeneous results but consistent evidence of the impact of IDEX on researchers’ networks.

2.2 Literature review and empirical framework

This paper speaks to at least three strands of literature. The first includes studies aiming to estimate the impact of public funding on scientific productivity. Although the vast majority of existing studies focus on competitive grants awarded to individual researchers (Arora and Gambardella, 2005; Jacob and Lefgren, 2011; Gush et al., 2018; Carayol and Lanoë, 2018; Heyard and Hottenrott, 2021), some studies focus on funding addressed to universities. The outcomes considered differ among studies, and the results are not always convergent. A stream of literature focuses on the change in the universities’ efficiency due to competitive funding. Johnes (2006, 2008) did not find efficiency improvements for universities in England, while Gawellek and Sunder (2016) documented an efficiency growth for German universities when awarded the *Exzellenzinitiative*, a competitive grant for excellence. A different stream of literature highlighted the unexpected negative outcomes of competitive funds addressed to universities. In Italy, Fadda et al. (2022) found discrimination in funding allocation in favor of universities already in a privileged position. Dougherty et al. (2016), used interviews to evidence that the limited impact of funding on student outcomes in the US is due to the increasing complexity of the internal organization. Closer to ours are studies focusing on the impact of US public funding on universities. Payne (2002) and Payne and Siow (2003), using instrumental variables for funding, found a positive effect of US research funding on universities’ publication productivity. They show that a \$1 million increase in funding generates 10–16 additional scientific articles and 0.2 more patents. Sacks (2007) investigated the effect of doubling the US

National Institutes of Health (NIH) budget between 1998 and 2002 on several US laboratories in biomedical fields. He did not find the expected jump in US researchers' productivity relative to a control sample of researchers affiliated with labs outside the US where funding did not double. Finally, Wahls (2018), also focusing on NIH funding, reported a positive effect of funding on universities' scientific output in terms of publications and citation productivity, but with a non-linear relation between funding amount and research outcome.

The second strand of literature relevant to our paper focuses on the impact of public funding on scientific outcomes related to the multiple activities performed by researchers. Emerging literature has shown a link between public funding and scientific collaborations (Adams et al., 2005; Lee and Bozeman, 2005; Wuchty et al., 2007). Defazio et al. (2009), analyzing the competitive EU *Framework Program*, found that collaborations created to capitalize on funding opportunities stimulate collaborations in the post-funding period that benefit researchers' productivity. Similarly, Davies et al. (2022) showed how the New Zealand competitive *Marsden Fund* increased by 13.8% the likelihood of researchers' collaboration post-funding if they had submitted a co-application together. Other than collaborations, governments can use funding instruments to foster researchers' interdisciplinary activities. The complexity reached by scientific research pushes to search for new knowledge through cross-disciplinary and cross-institutional collaborations (Geuna, 2001). Structural investments to universities are required to make them more suitable for this new process, and policymakers favor reorganizing universities into interdisciplinary poles (Gibbons et al., 1994; Rylance, 2015; Biancani et al., 2018; Hackett et al., 2021). Bruce et al., 2004, analyzing the case of the EU *Framework Program*, showed how overcoming barriers between disciplines and encouraging interdisciplinary collaboration through research funding is a difficult process.

Finally, literature shows that funding affects researchers' fundraising and patenting activities. Regarding fundraising, science is characterized by self-reinforcing mechanisms where granted researchers are more likely to accumulate additional grants in the future (Merton, 1968; Allison and Long, 1982; Bol et al., 2018). Concerning patenting, government funding seems less effective than industrial funding, or it is effective only in fostering applied research that is potentially patentable (Gulbrandsen and Smeby, 2005; Lawson, 2013; Hottenrott and Lawson, 2017). However, the increasing reliance on direct government funds for universities has led to increasing incentives for researchers to engage in technology transfer activities, and in France the propensity of researchers to invent has significantly increased in the last years (Geuna and Nesta, 2006; Carayol and Carpentier, 2021). Moreover, literature has not reported any trade-off for researchers between publishing and patenting (Breschi et al., 2007; Azoulay et al., 2009).

The third strand of literature to which this paper is related concerns the indirect effects of science. The small-world properties of the scientific collaboration networks are associated with researchers' publication productivity, team size, fundraising, and patenting (Eslami et al., 2013; Ebadi and Schiffauerova, 2015a and 2015b; Tahmooresnejad and Beaudry, 2019). Moreover, the connection with top-funded scholars is positively associated with researchers' number of publications (Mirnezami et al., 2020). Finally, researchers' industry engagement and patenting are influenced by peers' behaviors (Tartari et al., 2014; Carayol and Carpentier 2021).

2.2.1 *Empirical Framework: the IDEX program*

Several European countries have recently implemented structural reforms in higher education (De Boer et al., 2017). France launched its reform with the 2010 fiscal stimulus called the *Plan d'Investissements pour l'Avenir* (PIA) program[34], designed to respond to the economic crisis. Within PIA, a specific action addressed to universities was launched with the name of *Initiative D'EXcellence* (IDEX). With this program, the French government aimed to endow universities with more resources due to the absence of French universities in global rankings[35], the increased tension in the academic community asking for new investments[36], and some evidence on the lack of quality of French scientists' research outcomes (Enserink, 2008; Boudard and Westerheijden, 2017).

IDEX was designed as a direct government fund and used by the French government to distribute money to universities on a competitive basis with three main objectives, namely (i) supporting the selected universities to reach excellence at a global level, (ii) favoring the concentration of higher education and research institutions, and (iii) promoting economic growth in France. To achieve these objectives, IDEX provided financial resources to universities to renovate buildings and facilities, foster collaborations among complementary research entities geographically close to each other (universities, *Grandes Écoles*, and research institutions), increase universities' international competitiveness and attractiveness, promote interdisciplinary activities with the aim of creating 'multidisciplinary poles of excellence', create doctoral training programs of excellence[37], and foster researchers' patenting activity through the support of the SATT companies[38] (Legifrance, *Convention*

[34] In English, *Investments for the Future* program. PIA is divided into 35 actions grouped in five main axes: higher education and training (€11.9 billion), research (€7.9 billion), industry and SMEs (€6.5 billion), sustainable development (€5.1 billion), and digital (€4.5 billion).
[35] Also known as 'the Shanghai Shock', in reference to the Shanghai University Ranking of 2003.
[36] In 2004 the movement 'save the research' was created. Website: http://sauvonslarecherche.fr/ .
[37] See for example the IDEX Ph.D. program of Université Paris Saclay. Website: https://www.universite-paris-saclay.fr/en/universite-paris-saclay-international-joint-phd-program-cotutelle-2021.
[38] *Sociétés d'accélération du transfert de technologies*, in English Technology Transfer Acceleration Companies. These companies are directly connected to universities and researchers. See for example the *Aquitaine Science Transfert* SATT, linked to the University of Bordeaux. Website: https://www.ast-innovations.com/en/the-national-satt-network.

du 23 septembre 2010 entre l'Etat et l'ANR relative au programme d'investissements d'avenir, 2010; Finance Committee of the French National Assembly, Information report, 2015).

IDEX is divided into two rounds. The first round dates back to 2011, while the second round to 2015. This paper focuses on the first round since the second is currently ongoing[39]. The first round was endowed with a budget of €7.7 billion. Potential applicants of IDEX were the 26 *Higher Education and Research Clusters* resulting from the aggregation process between universities and the other geographically close research and academic entities (MESRI, 2019)[40]. Throughout the paper, we use the word 'university' to refer to a *HE&R Cluster* since the cluster's core is the university that is part of it (Boudard and Westerheijden, 2017)[41]. Table N1, in Appendix N, describes the 26 universities and their status concerning IDEX funding. Universities aspiring to IDEX were asked to submit a project proposal evaluated by an international jury based on four criteria: the quality of the teaching, the quality of the research, the connection with industry and the local sector, and the capacity of the university governance to manage the project. Seventeen universities applied for the first round of IDEX, and eight of them were awarded, namely Université d'Aix-Marseille, Université de Bordeaux, PSL Université Paris Sciences et Lettres, Université Paris Saclay, Sorbonne-Paris-Cité, Sorbonne Université, Université de Toulouse, and Université de Strasbourg. Awarded universities benefit from perpetual annual funding of about 25 million euros per year[42] (Legifrance, 2010; Finance Committee of the French National Assembly, Information report, 2015).

According to the goal of IDEX, we expect IDEX to affect four main areas of university researchers' activities: research, patenting, mentoring, and fundraising. Publications and collaborations are expected to be encouraged by the competitive mechanism of the funding allocation. Moreover, the geographical concentration of universities and research entities might favor economies of scale and scope that enhance research production (Geuna, 2001; Lepori et al., 2019). Public funding of universities might also increase researchers' probability of attracting other competitive grants due to the self-reinforcing mechanisms that occur in science. Finally, mentoring new Ph.D. students is an opportunity for researchers to create new research teams (Conti et al., 2014).

[39] Our period of analysis stops in 2015 to avoid overlapping with the second round.

[40] Composed of 19 communities of universities and institutions (COMUEs) and 7 associations of universities and institutions. We do not consider the 3 experimental public institutions because they were created during the second round of IDEX. Website: https://cache.media.enseignementsup-recherche.gouv.fr/file/Etablissements_et_organismes/68/2/Liste_regroupements_Associations_et_COMUE_et_associes_1er_fevrier_2018_890682.pdf.

[41] Most often, the name of the cluster coincides with the name of the university that is part of the cluster; for example, 'PSL Université Paris Sciences et Lettres'.

[42] Originally, IDEX was designed as a 4-year probationary period where universities were allowed to spend only the interests of the entire capital. After the probationary period, they would have received the entire capital. Since the entire capital was considered too large, it was later converted into a perpetual annual funding.

IDEX funding might also harm researchers' activities. IDEX might shrink university researchers' outcomes due to the loss in efficiency of universities during the application phase (Gawellek and Sunder, 2016). Moreover, outcomes can be hampered by the fact that competitive funds introduce constraints to universities' research lines and require reorganization efforts and additional time for adaptation to a new mechanism (Geuna, 2001; De Boer et al., 2017). We also expect a limited effect of IDEX on researchers' patenting activity due to the IDEX multi-goal nature that is likely to encourage researchers to pursue activities more aligned with traditional academic research (Perkmann et al., 2013). However, the competitive mechanism introduced by IDEX might stimulate the commercialization of university inventions (Goldfarb and Henrekson, 2003; Aghion et al., 2010).

We expect an impact also on researchers from universities that did not apply for IDEX but are connected with researchers who benefitted from IDEX funding through peer effects and knowledge flows within the collaboration networks.

2.3 Data

To assess the impact of the IDEX funding program on French researchers, we collect data from multiple sources. Information about French universities that applied and were awarded IDEX funding is drawn from the documentation provided by the French Minister of Higher Education, Research and Innovation (MESRI), and the General Investment Commission (*Commissariat Général à l'Investissement*).

Then, we collect French researchers' bibliometric records. We retrieve bibliometric data from Elsevier's SCOPUS database. Specifically, we collect all the STEM-related publications that report at least one researcher affiliated with a French university between 2002 and 2015. When selecting STEM fields, we refer to the *SCOPUS's All Science Journal Classification scheme*. We drop publications in Social Sciences, while we keep publications in Health Sciences, Life Sciences, and Physical Sciences. We end up with 1.28 million publications and 1.44 million authors.

We retrieve bibliometric information from the publications to construct appropriate bibliometric indicators for each researcher and university. Specifically, from each publication, we retrieve the authors' identifier, name and surname, the authors' research institutes of affiliation with information on the city and the country where the institutes are located, the type and the date of publication, the funding information, the journal of publication, and the stock of forward citations received by the publication at 2018. Moreover, we use SCOPUS to retrieve all the publications since 1990 of the authors in our final study sample, to proxy their seniority, and to assign the research fields to journals.

For each French researcher, we also construct a measure of the mentoring activity over time by identifying all the Ph.D. students supervised by the researcher during her career. To do so, we collect data on the whole universe of French Ph.D. thesis manuscripts in STEM disciplines, deposited from 2002 to 2015 in the French repository of *Electronic Doctoral Theses* (EDT). The EDT is the national centralized repository of French doctoral dissertations managed by the *Agence Bibliographique de l'Enseignement Supérieur* (ABES). For each thesis record, we have information on the author, the university and the city of graduation, the defense date, the supervisor's name, the co-supervisor's name (if any), and the field of study. We collect around 103 thousand Ph.D. theses.

To measure the patenting activity of French researchers, we gather patent data from the *European Patent Office* (EPO). Specifically, we retrieve all the French patent applications from 2002 to 2015. For each patent application, we collect data on the application's identifier and application date, inventors' identifier, name and surname, and inventors' residential address. We end up with around 135 thousand French patent applications and 114 thousand inventors.

Finally, we proxy the researchers' fundraising activity by relying on the ANR individual grants awarded to researchers. ANR grants are provided by the *French National Research Agency* (*ANR*), the most important French funding agency for research. ANR was founded in 2005 and aims to promote project-based research by distributing grants to researchers affiliated with French institutions. ANR awarded 1,157 research projects in 2019 with an average budget of 400 thousand euros per project. We use the complete list of individual grants awarded by ANR to French researchers between 2005 and 2015. We retrieve information on grants' identifier, the name and surname of the respective principal investigators, and the grants' starting date. We end up with around 12.9 thousand ANR grants.

2.3.1 Study Sample

To construct our study sample, we start by identifying all the researchers affiliated with a French university with at least one publication[43] covered by the SCOPUS database between 2006 and 2015.

We then restrict our sample by selecting only active researchers. We define a researcher as active in year *t* if she has at least one publication in the previous four years, i.e., from *t-4* to *t-1*. We select researchers who have been active in each of the ten years between 2006 and 2015. According to these criteria, the minimum number of publications that permit a researcher to be included in our sample is three publications between 2002 and 2014.

[43] We consider all type of publications to select our initial sample of researchers, i.e., we consider articles in peer-reviewed journals, books, conference papers, and reviews.

Then, we further restrict our sample by keeping only researchers whose affiliation has not changed during our analysis period[44]. We do so because researchers who changed affiliation over time might mix periods spent in IDEX universities with periods spent in non-IDEX universities, confounding the identification of the IDEX effect. For instance, a researcher who moves from a non-IDEX university to an IDEX university during the study period might be affected by IDEX only for a few years[45].

We rely on the affiliation reported in her SCOPUS publications to assign a researcher to a university. We assign a researcher to a university in year *t* if she has a publication in the previous 4 years, i.e., from *t-4* to *t-1*, reporting that university as affiliation.

Finally, we limit our sample to researchers in STEM fields. As STEM fields, we consider Health Sciences, Life Sciences, and Physical Sciences. We assign each researcher to a unique field according to the field of the journals where she published her research work during the period 2006-2015. Specifically, we identify the researcher's most frequent field over the ten years covered by our analysis, and we assign the researcher to that field.

We add to each researcher information on her patenting activity, fundraising, and Ph.D. students supervised in each of the ten years of our period of analysis. To add them, we use different approaches to minimize the incidence of incorrect matches based on researchers' names. When adding researchers' patent applications, we use two matching criteria: after selecting only French patent applications, we (i) match the first and last names of researchers to those of inventors who applied for patents, and we (ii) match the city of the researchers' university of affiliation retrieved from publications to the city of the inventors' residential address extracted from patent applications. When adding researchers' ANR grants, we use two other matching criteria: after obtaining the official list of ANR grants from the ANR agency, we (i) match the first and last names of researchers to those of the ANR Principal Investigators reported on the list, and we (ii) look at the researchers' publications whether they mention the ANR agency among the funding information in the acknowledgments. In half of the cases, the publications' funding information also reported the ANR grant number, allowing us to use it as a further matching criterion. Finally, to add information about the number of Ph.D. students supervised by the researchers, we use two matching criteria: we (i) match the first and last names of researchers to those of the Ph.D. students' supervisors reported on the thesis manuscripts, and we (ii) match the city of the researchers' university of affiliation to the city of the university

[44] We allow movement between universities having the same status with reference to the IDEX funding program. For instance, we keep a researcher who moved from Université Paris-Saclay to Sorbonne Université along our time window because both universities are IDEX awarded universities.

[45] Since we use a *Difference-in-differences* estimation approach to assess the IDEX effect, allowing for movements of researchers between universities risks to violate the SUTVA identification assumption (Rubin, 1980). In fact, movements may create interferences between researchears affected and non affected by IDEX. See section 2.5 for a detailed methodological explanation.

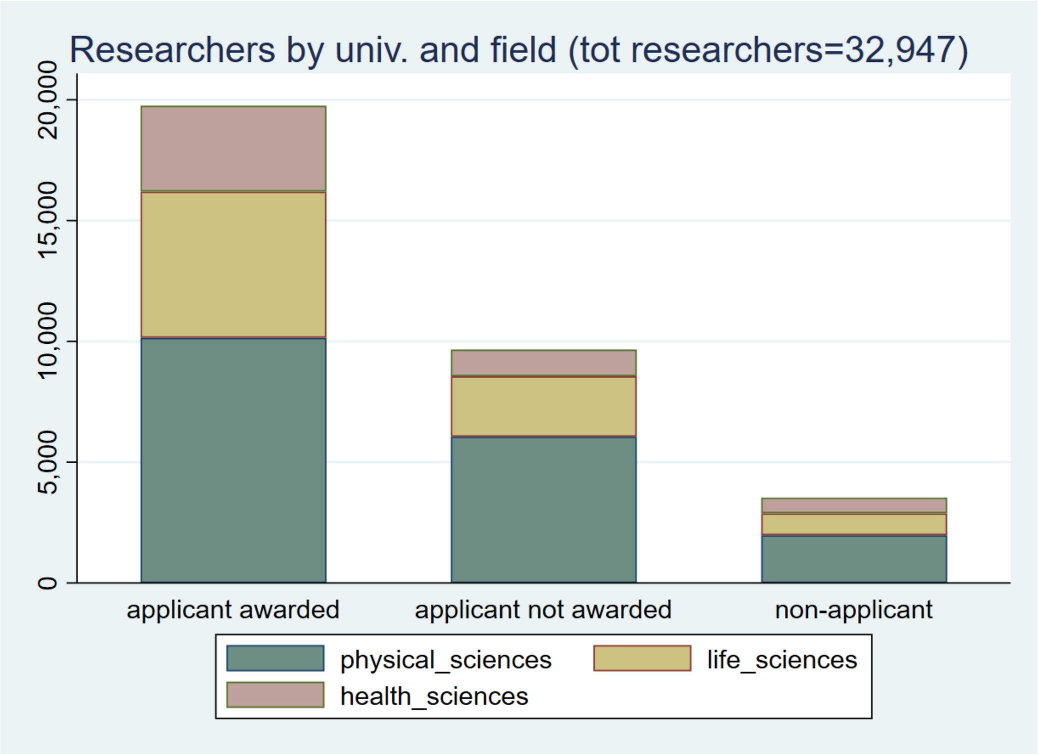
where the Ph.D. students defended their thesis. Our final goal is to identify the new Ph.D. students who start working with the researchers in our sample in each of the ten years of our period of analysis. Hence, we retrieve the starting date of the Ph.D. students assuming that each student started their Ph.D. program three years before the thesis defense date[46] (Corsini et al., 2022).

After matching, we obtain a study sample of 32,947 researchers active without interruptions over the ten years considered, i.e., between 2006 and 2015. Overall, our study sample is composed of a balanced panel dataset of 329,470 researcher-year pairs.

In our study, we focus on the impact of the first IDEX funding round that took place in 2011. In this round, eight universities out of seventeen applicant universities were awarded IDEX. Among the 32,947 researchers included in our study sample, 19,753 (60%) are affiliated with the eight universities awarded IDEX, while 9,660 (29%) are affiliated with the nine universities which applied but did not obtain the IDEX funding. The remaining 3,534 researchers (11%) are affiliated with the nine universities that did not apply for IDEX funding.

Figure 3 shows our study sample of researchers broken down by IDEX status of the university of affiliation and field.

Figure 3. Researchers by IDEX status of the university of affiliation and field.



[46] The Ph.D. students' thesis do not report the information on the exact starting date of the Ph.D., but they report the exact date of the final defense.

2.4 Variables

2.4.1 Dependent variables

We use as dependent variables a broad set of outcomes depicting the French researchers' activities.

We measure the quantity, quality, and interdisciplinarity of the researchers' publication outcomes in year *t* with the variables *Publications, Citation Weighted Publications,* and *Interdisciplinary Publications*. The variable *Publications* counts the number of scientific articles published by a researcher in peer-reviewed journals in year *t*[47]. The variable *Citation Weighted Publications* measures the quality-weighted publication productivity by weighting each article by the number of average forward citations received yearly. For instance, if a researcher publishes in year *t* two articles, one that received 0.5 average citations per year and the other 3 average citations per year, the variable *Citation Weighted Publications* takes in year *t* the value of 3.5 (0.5+3). Finally, we measure the interdisciplinary researcher productivity with the variable *Interdisciplinary Publications* which counts the number of articles published in peer-reviewed interdisciplinary journals in year *t*. We use the SCOPUS classification to identify interdisciplinary journals, which we define as those classified in more than one SCOPUS field and those classified as 'multidisciplinary' journals by SCOPUS. The latter category refers to generalist journals, such as *Nature* or *Science*.

Besides the researchers' publication productivity, we include a set of variables proxying the researchers' collaboration behaviors. We calculate the variable *Interdisciplinary Co-authors* that counts the number of distinct co-authors who belong to research fields different than that of the focal author in year *t*. We use a more granular classification of research fields to identify interdisciplinary co-authors than that used before for classifying STEM fields[48]. Due to data limitations, the variable *Interdisciplinary Co-authors* refers only to co-authors affiliated with French universities. Thus, we do not capture interdisciplinary collaborative behaviors that overcome the French borders. Nevertheless, the goal of IDEX to boost interdisciplinary is mainly focused on collaborations within France.

Moreover, we include four variables defining four different boundaries of the researchers' collaboration network. The variable *Within Lab Co-authors* counts the number of the researcher's distinct co-authors affiliated with the same laboratory as the focal researcher in year *t*. The variable *Within University Co-authors* counts the number of distinct co-authors affiliated with the same

[47] For the researchers' publication outcomes, we refer only to scientific articles published in peer-reviewed journals, i.e., excluding all the other type of publications.
[48] See Appendix M for a detailed explanation of the classification of research fields used for the variable *Interdisciplinary Co-authors*.

university as the focal researcher in year *t*, but in a different laboratory. The variable *National Co-authors* counts the number of the researcher's distinct co-authors affiliated with French universities, but outside the university of the focal researcher, in year *t*. Finally, the variable *International Co-authors* counts the number of researcher's distinct co-authors affiliated to a foreign, i.e., non-French, institution in year *t*. Defining the boundaries of researchers' collaborations allows us to explore the IDEX effect according to its objectives both to promote the geographical concentration of research institutes and to foster their internationalization.

Scientific productivity and collaborations are the main activities characterizing university researchers. Nevertheless, IDEX encourages other researchers' activities such as patenting, mentoring, and fundraising. Hence, we include three variables proxying the outcomes of these three activities. First, the dummy variable *At least one Patent* that takes value one if the researcher applies to at least one patent at the European Patent Office in year *t*, zero otherwise. Second, the dummy variable *At least one Ph.D. Student* that takes value one if the researcher starts mentoring at least one new Ph.D. student in year *t*, zero otherwise. Finally, the dummy variable *At least one ANR Grant* that takes value one if the researcher is awarded at least one ANR grant in year *t*, zero otherwise.

2.4.2 Main independent variables

We create three variables identifying the status of a researcher concerning IDEX funding based on her university of affiliation. Specifically, we calculate the variable *IDEX Applicant* as a dummy that equals one if the researcher's university of affiliation applied for IDEX funding, zero otherwise. Then, we calculate the variable *IDEX Awarded* as a dummy that equals one if the researcher's university of affiliation was awarded IDEX funding, zero otherwise.

Finally, for the sub-sample of researchers affiliated with universities that did not apply for IDEX (*Non-Applicants*), we calculate the variable *Connected to Awarded* in *t,* looking at the co-authorship network in the four years preceding *t*. The co-authorship network includes all the French researchers and their non-French co-authors. It is constructed considering only scientific articles with less than 20 authors to better proxy for actual collaborations between scientists (Gonzalez-Brambila et al., 2008). The variable *Connected to Awarded* is a dummy that equals one in year *t* if the researcher affiliated with a non-applicant university has a publication co-authored with a researcher affiliated with a IDEX awarded university in the four years preceding *t*, zero otherwise[49].

[49] Throughout the paper, we use the terms "awarded researcher", "applicant researcher", "applicant-non-award researcher", and "non-applicant researcher" to refer to researchers affiliated with universities awarded IDEX, universities that apply for IDEX, applicant-non-awarded universities, and non-applicant universities, respectively. We prefer to use this short version of the terminology to make the paper easier to read.

2.4.3 Controls

The researcher's academic achievements are affected by her past career results. In particular, senior scientists are more likely to have higher academic recognition that facilitates access to new resources and benefits their productivity, with higher gains from collaborative research (Merton, 1968; Allison and Stewart, 1974; Allison and Long, 1982; Gingras et al., 2008; Ebadi and Schiffaranova, 2015a).

Seniority influences researchers' outcomes independently from the impact of IDEX. Thus, we define the variable *Researcher's Seniority* in year t as the number of years elapsed between the researcher's first publication[50] and the year t (Nane et al., 2017). To capture possible nonlinear effects of seniority, we include a squared term of the seniority with the variable *Researcher's Seniority* ^{2}.

2.4.4 Descriptive statistics

Table 6 reports the descriptive statistics of the 32,947 researchers in our study sample observed between 2006 and 2015.

On average, a researcher in our sample publishes 2.36 peer-reviewed articles per year (8.21 when weighted by the yearly citations received and 1.27 when we consider interdisciplinary articles). The average researcher's network comprises 0.45 co-authors per year belonging to a different research field, 3.93 co-authors per year affiliated with the same lab, 0.49 co-authors per year affiliated with the same university but in a different lab, 1.49 national co-authors per year affiliated with a different French university, and 4.52 international co-authors per year. All the researchers in our sample have at least one publication during the ten years when they are observed, and on average, they have at least one publication in 73.4% of research-year observations. Over the ten years observed, an average researcher publishes 23.6 articles. The median is 16 articles.

Only 9.5% of researchers in our sample apply to at least one EPO patent between 2006 and 2015 (2.1% of research-year pairs). Almost half of our researchers, 47.2%, have mentored at least one new Ph.D. student during the ten years observed (12.9% of research-year pairs). Finally, 13.8% of researchers have been awarded at least one ANR grant between 2006 and 2015 (1.74% of research-year pairs). The researcher's average seniority is 14.45 years.

Researchers from awarded universities show higher publication productivity, in terms of quantity, quality, and interdisciplinarity, compared to the other groups of researchers in our sample. Furthermore, they have a larger network considering all four geographical proxies used: collaborations within the lab, collaborations within the university, national collaborations, and

[50] We retrieved data on researchers' publications since 1990.

international collaborations. They also show better outcomes when looking at the patenting and fundraising activities.

On the contrary, researchers from non-applicant universities score less than the other groups of researchers in our sample according to all the outcomes previously described. Finally, 30.7% of the researchers in non-applicant universities are co-authoring articles with researchers from awarded universities.

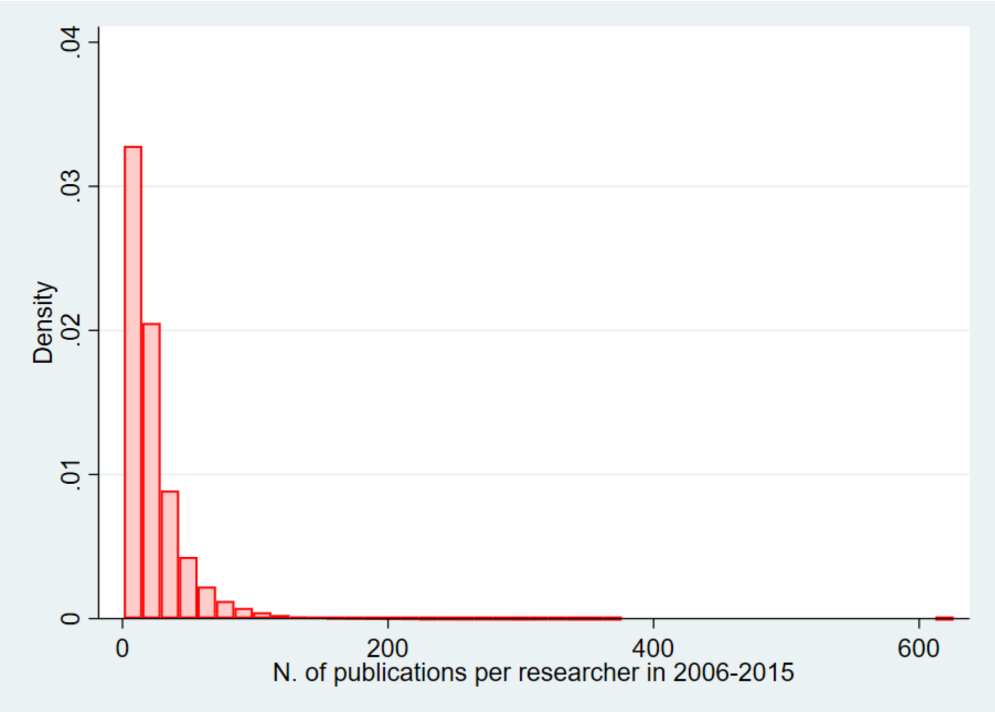
Figure 4 shows the distribution of the researchers' total number of publications over the ten years included in our analysis. Table 7 breaks down the descriptive statistics by research field.

Table 6. Descriptive statistics.

	Researchers	Observation Researcher-year	Mean outcome measure per year	sd	min	max	% zero values Researcher-year
Overall							
Publications	32,947	329,470	2.36	3.14	0	95	26.6%
Citation Weighted Publications	32,947	329,470	8.21	22.56	0	1815.8	27.9%
Interdisciplinary Publications	32,947	329,470	1.27	2.02	0	54	45.9%
Interdisciplinary Co-authors	32,947	329,470	0.45	0.70	0	5	64.5%
Within Lab Co-authors	32,947	329,470	3.93	5.84	0	117	40.1%
Within University Co-authors	32,947	329,470	0.49	1.75	0	59	85.2%
National Co-authors	32,947	329,470	1.49	4.15	0	142	69.9%
International Co-authors	32,947	329,470	4.52	15.84	0	812	57.9%
At least one Patent	32,947	329,470	0.021	0.14	0	1	97.9%
At least one Ph.D. Student	32,947	329,470	0.129	0.33	0	1	87.1%
At least one ANR Grant	32,947	329,470	0.017	0.13	0	1	98.2%
Researcher's Seniority	32,947	329,470	14.45	6.11	0	25	0.87%
Applicants							
Publications	29,413	294,130	2.41	3.21	0	95	26.2%
Citation Weighted Publications	29,413	294,130	8.60	23.51	0	1815.8	27.5%
Interdisciplinary Publications	29,413	294,130	1.30	2.07	0	54	45.5%
Interdisciplinary Co-authors	29,413	294,130	0.46	0.70	0	5	64.2%
Within Lab Co-authors	29,413	294,130	3.96	5.92	0	117	39.9%
Within University Co-authors	29,413	294,130	0.52	1.81	0	59	84.5%
National Co-authors	29,413	294,130	1.53	4.22	0	142	69.5%
International Co-authors	29,413	294,130	4.74	16.38	0	812	57.1%
At least one Patent	29,413	294,130	0.021	0.14	0	1	97.9%
At least one Ph.D. Student	29,413	294,130	0.131	0.34	0	1	86.9%
At least one ANR Grant	29,413	294,130	0.018	0.13	0	1	98.2%
Researcher's Seniority	29,413	294,130	14.46	6.12	0	25	0.77%
Awarded							
Publications	19,753	197,530	2.52	3.41	0	95	25.7%
Citation Weighted Publications	19,753	197,530	9.71	26.44	0	1815.8	26.9%
Interdisciplinary Publications	19,753	197,530	1.33	2.20	0	54	45.6%
Interdisciplinary Co-authors	19,753	197,530	0.46	0.69	0	5	64.0%
Within Lab Co-authors	19,753	197,530	4.25	6.43	0	117	39.4%
Within University Co-authors	19,753	197,530	0.58	1.99	0	59	83.6%
National Co-authors	19,753	197,530	1.80	4.71	0	142	66.7%

International Co-authors	19,753	197,530	5.54	18.45	0	812	56.1%
At least one Patent	19,753	197,530	0.022	0.15	0	1	97.8%
At least one Ph.D. Student	19,753	197,530	0.121	0.33	0	1	87.9%
At least one ANR Grant	19,753	197,530	0.019	0.14	0	1	98.1%
Researcher's Seniority	19,753	197,530	14.55	6.14	0	25	0.50%
Non-applicants							
Publications	3,534	35,340	1.93	2.43	0	83	29.6%
Citation Weighted Publications	3,534	35,340	5.02	11.6	0	487.14	31.4%
Interdisciplinary Publications	3,534	35,340	1.05	1.6	0	27	49.7%
Interdisciplinary Co-authors	3,534	35,340	0.43	0.67	0	5	66.3%
Within Lab Co-authors	3,534	35,340	3.65	5.1	0	82	41.3%
Within University Co-authors	3,534	35,340	0.25	1.14	0	29	91.2%
National Co-authors	3,534	35,340	1.16	3.42	0	80	73.6%
International Co-authors	3,534	35,340	2.73	10.22	0	369	64.1%
At least one Patent	3,534	35,340	0.017	0.13	0	1	98.3%
At least one Ph.D. Student	3,534	35,340	0.118	0.32	0	1	88.2%
At least one ANR Grant	3,534	35,340	0.011	0.11	0	1	98.9%
Researcher's Seniority	3,534	35,340	14.41	6.05	0	25	0.10%
Connected to awarded	3,534	35,340	0.307	0.46	0	1	69.3%

Figure 4. Distribution of the researchers' total number of publications in 2006-2015.



NOTE: 5% of researchers have more than 68 publications. The researcher with the largest number of publications (626) is Raoult Didier, an infectivologist affiliated with Université d'Aix-Marseille, Marseille, France, since 1986.

Table 7. Descriptive statistics by field.

	Health Sciences (5,308 researchers)	Life Sciences (9,465 researchers)	Physical Sciences (18,174 researchers)
Overall			
Observations	53,080	94,650	181,740
Publications	3.65	1.96	2.19
Citation Weighted Publications	14.11	7.77	6.72
Interdisciplinary Publications	1.40	1.13	1.31
Interdisciplinary Co-authors	0.70	0.51	0.35
Within Lab Co-authors	6.37	4.41	2.96
Within University Co-authors	0.61	0.50	0.45
National Co-authors	2.09	1.51	1.31
International Co-authors	5.81	3.56	4.64
At least one Patent	0.017	0.023	0.020
At least one Ph.D. Student	0.069	0.108	0.157
At least one ANR Grant	0.007	0.022	0.018
Researcher's Seniority	16.20	15.92	14.98
Applicants			
Observations	46,580	85,570	161,980
Publications	3.82	1.99	2.23
Citation Weighted Publications	15.16	8.04	7.00
Interdisciplinary Publications	1.47	1.14	1.33
Interdisciplinary Co-authors	0.73	0.51	0.35
Within Lab Co-authors	6.58	4.40	2.97
Within University Co-authors	0.62	0.53	0.48
National Co-authors	2.18	1.55	1.34
International Co-authors	6.30	3.71	4.83
At least one Patent	0.019	0.024	0.020
At least one Ph.D. Student	0.072	0.108	0.158
At least one ANR Grant	0.008	0.023	0.018
Researcher's Seniority	16.25	15.93	14.97
Awarded			
Observations	35,530	60,530	101,470
Publications	3.98	2.00	2.32
Citation Weighted Publications	16.43	8.60	8.03
Interdisciplinary Publications	1.55	1.14	1.37
Interdisciplinary Co-authors	0.72	0.51	0.33
Within Lab Co-authors	7.17	4.50	3.08
Within University Co-authors	0.71	0.55	0.56
National Co-authors	2.49	1.73	1.60
International Co-authors	6.82	4.02	5.99
At least one Patent	0.019	0.026	0.020
At least one Ph.D. Student	0.068	0.106	0.149
At least one ANR Grant	0.008	0.025	0.019
Researcher's Seniority	16.26	15.91	15.07
Non-applicants			
Observations	6,500	9,080	19,760
Publications	2.42	1.75	1.85
Citation Weighted Publications	6.58	5.18	4.44
Interdisciplinary Publications	0.89	1.04	1.11
Interdisciplinary Co-authors	0.54	0.50	0.35

Within Lab Co-authors	4.85	4.53	2.86
Within University Co-authors	0.52	0.30	0.14
National Co-authors	1.43	1.18	1.05
International Co-authors	2.31	2.15	3.14
At least one Patent	0.008	0.022	0.018
At least one Ph.D. Student	0.047	0.100	0.150
At least one ANR Grant	0.004	0.014	0.013
Researcher's Seniority	15.84	15.83	15.06
Connected to awarded	0.284	0.335	0.301

2.5 Econometric methodology

Our analysis is in two steps and aims to assess (i) the direct effect of applying and being awarded IDEX and (ii) the indirect effect on non-applicant researchers of collaborating with awarded researchers. For this purpose, we conduct two separate empirical analyses using a *Difference-in-differences* approach (Angrist and Pischke, 2009). The first analysis considers the sample of all potential applicants, i.e., the 32,947 active researchers affiliated with French universities during the ten years considered, while the second considers the sample of 3,534 non-applicant researchers.

In the model presented in Equation 2, we evaluate the impact of applying and being awarded IDEX for the 32,947 active researchers. On the left-hand side of the equation, we consider the researcher *i*'s outcome as measured by one of the eleven variables described in section 2.4.1, i.e., *Publications*, *Citation Weighted Publications*, *Interdisciplinary Publications*, *Interdisciplinary Co-authors*, *Within Lab Co-authors, Within University Co-authors*, *National Co-authors*, *International Co-authors*, *At least one Patent*, *At least one Ph.D. Student*, *At least one ANR Grant.* On the right-hand side of the equation, we consider the dummy variables *IDEX Applicant* and *IDEX Awarded*, the control variable *Researcher's Seniority*, along with its squared term, and the dummy variable *After 2011* that equals one if *t* is greater than or equal to 2011. We consider the year 2011 as the first year when the awarded universities are expected to show the effect of IDEX. To estimate the effect of being affiliated with an applicant university after the application time, we interact the variable *IDEX Applicant* with the dummy variable *After 2011*. To estimate the effect of being affiliated with an awarded university after the IDEX award, we interact the variable *IDEX Awarded* with the dummy variable *After 2011*. We add to the model the year fixed effects (γ_t) and the researcher fixed effects (α_i), and cluster the standard errors around the researcher[51]. Researcher fixed effects allow us to control for researchers' time-invariant unobserved characteristics, such as inner ability and

[51] A difference-in-differences exercise requires to include in the equation the non-interacted dummy variables *IDEX Applicant, IDEX Awarded*, and *After 2011*. However, these dummies are collinear to the researcher and year fixed effects and are not included in the model of Equation 2.

motivation, that predetermine differences among researchers[52]. Year-fixed effects allow us to control for time effects and trends in the outcomes. In Equation 2, the effect of applying is estimated by β_1. Conditional on applying, the additional effect of being awarded IDEX is estimated by β_2[53]. Finally, $\varepsilon_{i,t}$ is the idiosyncratic error term. We estimate the coefficients of the model presented in Equation 2 using *Ordinary Least Squares* (OLS)[54]. When the dependent variable is binary, we estimate a *Linear Probability Model* (LPM)[55]. The analysis is at the researcher-year level.

$$\begin{aligned} Researcher\ outcome_{i,t} \\ = \beta_0 + \beta_1\ After\ 2011 * IDEX\ Applicant_i + \beta_2\ After\ 2011 * IDEX\ Awarded_i \\ + \beta_3\ Researcher's\ Senioriy_{i,t} + \beta_4\ Researcher's\ Senioriy^2{}_{i,t} + \gamma_t + \alpha_i + \varepsilon_{i,t} \end{aligned}$$

Equation 2

The model in Equation 3 estimates the indirect effect on the 3,534 non-applicant researchers collaborating with the awarded researchers. As for Equation 2, the dependent variable takes, in turn, the values of the eleven researchers' outcomes presented in section 2.4.1. On the right-hand side of the equation, we consider the dummy variable *Connected to awarded*, the control variable *Researcher's Seniority*, along with its squared term, and the dummy variable *After 2011* that equals one if *t* is greater than or equal to 2011. Similarly to Equation 2, the effect of collaborating with an awarded researcher after the IDEX award is estimated by the coefficient β_1, namely, the coefficient of the interaction between the variable *Connected to awarded* and the dummy *After 2011*. We include in Equation 3 the year fixed effects (γ_t) and the researcher fixed effects (α_i), and cluster the standard errors around the researcher[56]. Finally, $\varepsilon_{i,t}$ is the idiosyncratic error term. As for Equation 2, we estimate the coefficients of the model presented in Equation 3 using *Ordinary Least Squares* (OLS)[57]. When the dependent variable is binary, we estimate a *Linear Probability Model* (LPM). The analysis is at the researcher-year level.

---

[52] In an alternative model specification, we use university fixed effects and clustered standard errors around the university. Our results are unchanged and available upon request.

[53] All the awarded researchers are also applicant researchers. Thus, β_2 estimates the additional effect of being awarded IDEX beyond the application effect. Doing so, we can disentangle the effect of applying from the effect of being awarded IDEX.

[54] In section 2.8.4, we propose a robustness check where we use the Poisson pseudo-maximum likelihood estimator when the dependent variable is continuous.

[55] A Linear Probability Model might not be the best estimator in the case of binary outcome variables. However, the coefficients of the interaction terms in nonlinear models using a difference-in-differences method are not readily interpretable (Karaca-Mandic et al., 2012).

[56] Differently from Equation 2, Equation 3 reports the non-interacted variable *Connected to awarded* since, being this variable time-variant, it is not collinear with the researcher fixed effects. As for Equation 2, the non-interacted variable *After 2011* is collinear to the year fixed effects and is not included in the model of Equation 3.

[57] In section 2.8.4, we propose a robustness check where we use the Poisson pseudo-maximum likelihood estimator when the dependent variable is continuous.

$$Researcher\ outcome_{i,t} = \beta_0 + \beta_1\ After\ 2011 * Connected\ to\ awarded_{i,t} + \beta_2\ Connected\ to\ awarded_{i,t} + \beta_3\ Researcher's\ Seniority_{i,t} + \beta_4\ Researcher's\ Seniority^2{}_{i,t} + \gamma_t + \alpha_i + \varepsilon_{i,t}$$

Equation 3

Evaluating the relationship between IDEX funding and researchers' outcomes might cause an endogeneity problem due to the possible correlation between the unobserved characteristics of the researcher and the probability of her university to be awarded IDEX funding. For instance, public funds are preferentially assigned to outstanding researchers expected to devote part of their time to collaborative activities such as team-based goals, projects, and publications (Jaffe, 2002; Lee and Bozeman, 2005). Thus, universities recruiting researchers with higher ability and better outcomes might be more likely to be awarded funding from the French government. At the same time, the researchers affiliated with these universities are those expected to be more productive according to their characteristics. This scenario leads to a possible selection bias. Nonetheless, the IDEX funding program is the ideal empirical framework to mitigate this endogeneity problem for four main reasons. First, IDEX is assigned to universities based on four criteria, i.e., the quality of the teaching, the quality of the research, the connection with industry and the local sector, and the capacity of the university governance to manage the project. The evaluation of universities is made in a comprehensive way, and it is only partially connected to the researchers' outcomes. For instance, the reliability of university governance plays a crucial role. Université Paris Saclay and Université de Toulouse struggled in obtaining IDEX because they did not evince a solid governance, despite the high quality of their research. The application of Languedoc-Roussillon Universités (Montpellier) was rejected for the same reason (Finance Committee of the French National Assembly, Information report, 2015). Second, IDEX funding is awarded to universities, while our analysis is conducted at the researcher-year level. The correlation between the single researcher's characteristics and the public institution's probability of being awarded is likely to be limited. Third, we include in the regression exercises the researcher fixed effects to account for all the unobserved time-invariant characteristics of French researchers, such as inner ability and motivation. Fourth, in section 2.8.1 we propose two robustness checks where, first, we include in the regressions a set of time-variant university characteristics to control for observed factors that vary over time and may relate to the IDEX endowment and, at the same time, affect the researchers' outcomes of our treated and control groups differently. Specifically, we include proxies for the university size, quality, and fundraising. Second, we rely on the *Coarsened Exact Matching* (CEM) to create a control group of applicant-non-award researchers similar to the group of awarded researchers, controlling for differential outcome trends and stocks before IDEX. Including time-variant controls and CEM matching are expected to mitigate the possible estimation bias.

Finally, we ensure that the assumptions that make the *Difference-in-differences* approach valid for inference are satisfied. Specifically, we test for the key identifying assumption of the *Difference-in-differences* approach, namely the common trends assumption (Angrist and Pischke, 2009). This assumption implies that trends in researchers' outcomes would be the same in the absence of IDEX. We test for this assumption by employing a regression based test augmenting the base model of Equation 2 with the leads and lags, following Autor (2003) and Angrist and Pischke (2009). The test and the discussion of the common trends assumption in our study sample are reported in Appendix L. Concerning the stable unit treatment value assumption (SUTVA) (Rubin, 1980), we rely on the fact that in our study sample we keep only researchers whose affiliation has not changed during our analysis period. In doing so, we aim to avoid interferences between researchers affected and non-affected by IDEX, preventing IDEX treatment from influencing the potential outcomes of researchers in non-IDEX universities, and vice versa. Moreover, clustering the standard errors at the level of the researcher allows us to mitigate the correlation between researchers. Finally, since researchers outside our study sample may still move between universities and thus indirectly influence our study sample's researchers, we rely on the robustness check proposed in section 2.8.1. In this robustness check, we control for the time-variant characteristics of the universities, considering their yearly quality, size, and funding attractiveness. In calculating these variables, we consider all the university affiliate researchers each year, including also those who have changed affiliation during our study period.

2.6 Results

Table 8 shows the OLS estimates of the model described in Equation 2, which defines the direct effect of applying and being awarded IDEX funding on French researchers' activities. Columns 1, 2, and 3 show the IDEX effect on the French researchers' publication productivity. Columns 4, 5, 6, 7, and 8 show the IDEX effect on the researchers' collaboration behaviors. Finally, columns 9, 10, and 11 show the IDEX effect on the researchers' patenting, mentoring, and fundraising activities.

We find a significant effect of both being affiliated with a university that applies for IDEX and being affiliated with a university that is awarded IDEX funding on researchers' activities. This finding suggests that the university application process aimed at obtaining IDEX funding impacts the affiliate researchers, regardless of the result of the application. Specifically, we find that researchers affiliated with universities that applied for IDEX but did not obtain IDEX funding increase the number of interdisciplinary co-authors by 0.019 co-authors per year, the number of co-authors within the same laboratory by 0.14 co-authors per year, and the number of co-authors within the same university (but outside the lab) by 0.10 co-authors per year, relative to researchers in universities that did not apply for IDEX funding.

Moreover, if the university is awarded IDEX funding, affiliate researchers gain additional benefits. The effect of being awarded IDEX is calculated as the sum of the coefficients of the variables *After 2011*IDEX Applicant* and *After 2011*IDEX Awarded*. Specifically, we find that awarded researchers increase the number of interdisciplinary publications by 0.041 articles per year, the number of co-authors within the lab by 0.088 co-authors per year, the number of co-authors within the university (but outside the lab) by 0.090 co-authors per year, the number of national co-authors (but outside their university) by 0.35 co-authors per year, and the number of international co-authors by 1.06 co-authors per year, and they experience a decrease in the number of interdisciplinary co-authors by 0.020[58], relative to researchers in applicant universities that did not obtain IDEX funding.

Considering the magnitude of the estimated coefficients, the direct impact of IDEX funding on French researchers appears to be relevant to researchers' networks: IDEX application and the IDEX awarding increase researchers' collaborations within their university (both within and outside their lab). Moreover, when the university is awarded IDEX, researchers gain new national co-authors (outside their university) and about one new international co-author per year relative to researchers affiliated with applicant but not non-awarded universities. We do not find any significant effects of applying or being awarded IDEX on researchers' publication quality and quantity, patenting, mentoring, and fundraising.

[58] This effect compensates the positive effect obtained in the application phase. The linear combination of the coefficients of the variables *After 2011*IDEX Applicant* and *After 2011*IDEX Awarded* is not statistically different from zero (Pvalue=0.890). We conducted an F-test on the null hypothesis that β_1(*After 2011* IDEX Applicant*) + β_2(*After 2011*IDEX Awarded*) = 0.

Table 8. Direct effect of applying and being awarded IDEX on French researchers' outcomes. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*IDEX Applicant	**0.040**	**-0.0100**	**0.014**	**0.019****	**0.14****	**0.10*****	**-0.055**	**0.13**	**0.0017**	**0.0045**	**-0.0021**
	(0.034)	**(0.17)**	**(0.020)**	**(0.0082)**	**(0.064)**	**(0.016)**	**(0.044)**	**(0.13)**	**(0.0016)**	**(0.0036)**	**(0.0013)**
After 2011*IDEX Awarded	**0.033**	**0.21**	**0.041*****	**-0.020*****	**0.088****	**0.090*****	**0.35*****	**1.06*****	**-0.0011**	**0.0032**	**-0.0010**
	(0.023)	**(0.15)**	**(0.014)**	**(0.0053)**	**(0.042)**	**(0.013)**	**(0.031)**	**(0.10)**	**(0.0011)**	**(0.0024)**	**(0.00095)**
Res. Seniority	0.11***	0.49***	0.080***	0.0035	0.096***	0.050***	0.15***	0.26***	0.00091	0.020***	0.0064***
	(0.011)	(0.068)	(0.0081)	(0.0050)	(0.033)	(0.0083)	(0.019)	(0.062)	(0.0013)	(0.0041)	(0.0014)
Res. Seniority ^2	-0.0044***	-0.012***	-0.0023***	-0.00073***	-0.0030***	-0.00043***	0.00022	-0.00069	-0.000044***	-0.00063***	-0.00010***
	(0.00018)	(0.0013)	(0.00010)	(0.000038)	(0.00034)	(0.00010)	(0.00026)	(0.00091)	(0.000)	(0.000016)	(0.000)
Constant	1.56***	3.50***	0.55***	0.45***	2.49***	-0.17*	-0.92***	0.44	0.013	0.0091	-0.032**
	(0.11)	(0.64)	(0.083)	(0.053)	(0.34)	(0.085)	(0.19)	(0.61)	(0.014)	(0.044)	(0.015)
Observations	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470
R-squared	0.009	0.002	0.010	0.008	0.012	0.005	0.016	0.007	0.000	0.008	0.001
Number of researchers	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample of active French researchers, including 329,470 researcher-year pairs, 32,947 researchers observed for ten years each. Standard errors are clustered around the researcher. Standard errors are reported in parentheses. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variables *IDEX Applicant, IDEX Awarded,* and *After 2011.* However, these dummies are collinear to the researcher and year fixed effects and are not included in the regressions.

Table 9 reports the OLS estimates of the model described in Equation 3 that assesses the indirect impact of IDEX on the outcomes of researchers affiliated with non-applicant universities. We aim to estimate the possible spillover effects of IDEX funding on researchers who did not directly benefit from IDEX but are connected to awarded researchers through the co-authorship network.

We find a positive effect of IDEX funding on non-applicant researchers regarding the increase of their scientific network. Table 9 shows that non-applicant researchers who are connected to researchers awarded IDEX increase the number of co-authors within the same laboratory by 0.42 co-authors per year, the number of national co-authors (outside their university) by 0.30 co-authors per year, and the number of international co-authors by 0.94 co-authors per year, relative to non-applicant researchers who are not connected to researchers awarded IDEX. We do not find any significant effect on the other non-applicant researchers' outcomes.

Table 9. Indirect effect of collaborating with IDEX awarded for non-applicant researchers. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*Connect. to Awarded	**-0.0032**	**-0.53**	**-0.00054**	**0.011**	**0.42*** **	**-0.050**	**0.30*** **	**0.94*** **	**0.0046**	**-0.0037**	**0.0013**
	(0.069)	**(0.33)**	**(0.042)**	**(0.016)**	**(0.13)**	**(0.034)**	**(0.097)**	**(0.28)**	**(0.0036)**	**(0.0075)**	**(0.0027)**
Connected to awarded	0.070	0.17	0.041	-0.0088	-0.047	0.046*	-0.11	-0.064	0.00085	0.0074	-0.0027
	(0.048)	(0.23)	(0.031)	(0.012)	(0.097)	(0.027)	(0.076)	(0.17)	(0.0026)	(0.0063)	(0.0022)
Res. Seniority	0.079***	0.28**	0.087***	0.0079	0.14**	-0.0086	0.10***	0.36***	0.0037***	0.034***	0.0015
	(0.027)	(0.11)	(0.019)	(0.010)	(0.060)	(0.014)	(0.037)	(0.13)	(0.0011)	(0.012)	(0.0011)
Res. Seniority ^2	-0.0034***	-0.0053**	-0.0022***	-0.00049***	-0.0031***	0.00020	0.00052	-0.0049**	-0.000069***	-0.00053***	-0.000050***
	(0.00053)	(0.0025)	(0.00028)	(0.00011)	(0.00094)	(0.00022)	(0.00065)	(0.0023)	(0.000021)	(0.000043)	(0.000016)
Constant	1.41***	2.20**	0.32*	0.34***	1.92***	0.24*	-0.47	-1.09	-0.015	-0.18	0.0042
	(0.24)	(0.96)	(0.19)	(0.10)	(0.58)	(0.13)	(0.33)	(1.12)	(0.010)	(0.13)	(0.011)
Observations	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340
R-squared	0.006	0.002	0.010	0.007	0.009	0.001	0.010	0.004	0.001	0.009	0.001
Number of researchers	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample composed of 35,340 non-applicant researcher-year pairs, 3,534 researchers observed for ten years each. Standard errors are clustered around the researcher. Standard errors are reported in parentheses. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variable *After 2011*. However, this dummy is collinear to the year fixed effects and is not included in the regressions.

2.7 Field heterogeneity

In this section, we explore field heterogeneity of the direct and indirect effects of IDEX funding on French researchers' outcomes by conducting separate analyses by field. We report the results of our regression exercises for three main categories of STEM disciplines: Health Sciences, Life Sciences, and Physical Sciences.

Table 10 reports the OLS estimates of the direct impact of IDEX by field of research, while Table 11 reports the OLS estimates of the indirect impact of IDEX by field of research.

Table 10 shows how IDEX has been particularly beneficial for French researchers in the field of Health Sciences. Researchers affiliated with universities that applied for IDEX but did not obtain IDEX increase their publication productivity by 0.29 articles per year, 1.22 citation-weighted articles per year, and 0.12 interdisciplinary publications per year. Moreover, they enlarge their collaboration network by gaining 0.059 new interdisciplinary co-authors per year, 0.91 new co-authors per year within the laboratory, 0.19 new co-authors per year within their same university (but outside the lab), and 0.74 new international co-authors per year, relative to researchers in universities that did not apply for IDEX funding. Then, if the university of affiliation is awarded IDEX, researchers in Health Sciences further boost their productivity. They publish 0.089 more interdisciplinary articles per year, and they gain 0.16 new co-authors per year within their same university (but outside the lab), 0.85 new national co-authors per year (outside their university), and 2.27 new international co-authors per year, relative to researchers in universities that applied for IDEX but did not get IDEX funding. They also experience a reduction in the number of interdisciplinary co-authors per year by 0.033[59] and in the probability of obtaining an ANR grant by 0.31 percentage points. The effect of obtaining IDEX funding on researchers in Health Sciences is relevant for internationalization: the number of international co-authors increases by around three co-authors per year relative to researchers in universities that did not apply for IDEX[60].

Looking at the direct effect of IDEX on researchers in Life Sciences, we find that they increase the number of co-authors within the same university (but outside the lab) by 0.13 co-authors per year and the international co-authors by 0.39 co-authors per year when applying for IDEX funding, relative to researchers from universities that did not apply for IDEX. Then, if the university is awarded IDEX funding, affiliate researchers in Life Sciences increase their number of interdisciplinary publications

[59] This effect compensates the positive effect obtained in the application phase. The linear combination of the coefficients of the variables *After 2011*IDEX Applicant* and *After 2011*IDEX Awarded* is not statistically different from zero (Pvalue=0.130). We conducted an F-test on the null hypothesis that β_1(*After 2011* IDEX Applicant*) + β_2(*After 2011*IDEX Awarded*) = 0.

[60] The effect of being awarded IDEX relative to researchers in non-applicant universities is the linear combination of the coefficients of the variables *After 2011*IDEX Applicant* and *After 2011*IDEX Awarded*. Looking at Table 10, it is the linear combination of the estimates in the first two rows.

by 0.044 articles per year and further enlarge their collaboration network by gaining 0.047 co-authors per year within university (but outside the lab), 0.17 national co-authors per year (outside their university), and 0.55 international co-authors per year, relative to researchers in universities that applied for IDEX but did not get IDEX funding. When awarded, they also experience a slight decrease in the number of interdisciplinary collaborations by 0.018 co-authors per year and in the probability of obtaining an ANR grant by 0.33 percentage points. Also in this case, the effect of obtaining IDEX funding is relevant to what concerns international co-authors. Awarded researchers in Life Sciences gained around one new international co-author relative to researchers in universities that did not apply for IDEX[61].

Finally, we find that researchers in Physical Sciences benefit from IDEX for what concerns the collaboration networks and the patenting and mentoring activities. If their university applied for IDEX funding, they enlarge their network by gaining 0.18 new co-authors per year within the laboratory and 0.079 new co-authors per year within their university (but outside the lab), and increase the probability of applying for a patent by 0.37 percentage points and the probability of supervising a new Ph.D. student by 1.3 percentage points, relative to researchers in universities that did not apply for IDEX. If the university is awarded IDEX, affiliate researchers further extend their network by gaining 0.062 new co-authors per year within their university (but outside the lab), 0.16 new national co-authors per year (outside their university), and 0.70 new international co-authors per year, but decreasing the number of interdisciplinary co-authors by 0.023 co-authors per year, relative to researchers in applicant but not awarded universities.

Table 11 shows the results of the estimation of the indirect effect of IDEX on researchers affiliated with non-applicant universities, by field of research. Researchers in Health Sciences and Life Sciences in non-applicant universities increase their number of national co-authors (outside their university) by 0.75 and 0.43 co-authors per year, respectively, if connected with IDEX awarded researchers. Researchers in Life Sciences also increase the number of interdisciplinary co-authors per year by 0.065 if connected with awarded researchers. Instead, non-applicant researchers in Physical Sciences increase the number of co-authors within the laboratory by 0.42 co-authors per year and the number of international co-authors by 1.22 co-authors per year, and decrease the number of co-authors within their university (but outside the lab) by 0.082 co-authors per year, relative to researchers in non-applicant universities who are not connected to awarded researchers.

Researchers in Life Sciences are the only ones who experience indirect effects on publication productivity. Specifically, they decrease the number of citation-weighted publications by 0.95 if

[61] The effect of being awarded IDEX relative to researchers in non-applicant universities is the linear combination of the coefficients of the variables *After 2011*IDEX Applicant* and *After 2011*IDEX Awarded*.

connected to awarded researchers, relative to non-applicants who are not connected to awarded researchers.

Table 10. Direct effect of applying and being awarded IDEX on French researchers' outcomes, by field. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
HEALTH SCIENCES	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*IDEX Applicant	**0.29*** **	**1.22***	**0.12*** **	**0.059*** **	**0.91*** **	**0.19*** **	**-0.11**	**0.74***	**-0.00085**	**-0.0071**	**-0.0017**
After 2011*IDEX Awarded	**0.034**	**0.24**	**0.089** **	**-0.033** **	**0.034**	**0.16*** **	**0.85*** **	**2.27*** **	**0.00065**	**0.0067**	**-0.0031***
Res. Seniority	0.17*	1.38**	0.096*	-0.062	-0.042	0.14***	0.53***	0.85***	-0.0096	0.026***	-0.0017
Res. Seniority ^2	-0.0049***	-0.015***	-0.0017***	-0.00083***	-0.0034***	-0.00027	0.0032***	0.010***	-0.000040**	-0.00035***	-0.000046***
Constant	1.92*	-2.52	0.26	1.48***	5.87*	-1.25***	-5.92***	-8.51***	0.13	-0.17***	0.039
Observations	53,080	53,080	53,080	53,080	53,080	53,080	53,080	53,080	53,080	53,080	53,080
R-squared	0.016	0.004	0.016	0.009	0.018	0.013	0.057	0.030	0.001	0.004	0.001
Number of researchers	5,308	5,308	5,308	5,308	5,308	5,308	5,308	5,308	5,308	5,308	5,308
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LIFE SCIENCES											
After 2011*IDEX Applicant	**-0.0094**	**-0.19**	**-0.025**	**0.0096**	**-0.23**	**0.13*** **	**-0.077**	**0.39** **	**-0.0021**	**-0.0030**	**-0.0024**
After 2011*IDEX Awarded	**0.022**	**-0.091**	**0.044***	**-0.018***	**0.057**	**0.047***	**0.17*** **	**0.55*** **	**-0.0018**	**-0.0032**	**-0.0033***
Res. Seniority	0.030	0.11	0.078**	-0.049	-0.15	-0.019	0.27***	0.38***	0.0030***	0.030***	0.012*
Res. Seniority ^2	-0.0036***	-0.013***	-0.0018***	-0.00070***	-0.0022***	-0.00049***	-0.00063	-0.0021	-0.000011	-0.00063***	-0.00015***
Constant	2.03***	7.98**	0.32	1.10***	5.36***	0.58	-2.20***	-1.62	-0.014	-0.12***	-0.082
Observations	94,650	94,650	94,650	94,650	94,650	94,650	94,650	94,650	94,650	94,650	94,650
R-squared	0.009	0.004	0.011	0.006	0.016	0.007	0.017	0.008	0.000	0.008	0.003
Number of researchers	9,465	9,465	9,465	9,465	9,465	9,465	9,465	9,465	9,465	9,465	9,465
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

PHYSICAL SCIENCES	(1) Pubs	(2) Citation Weighted Pubs	(3) Interdisciplinary Pubs	(4) Interdisciplinary Co-authors	(5) Within Lab Co-authors	(6) Within University Co-authors	(7) National Co-authors	(8) International Co-authors	(9) At least one Patent	(10) At least one Ph.D. Student	(11) At least one ANR Grant
After 2011*IDEX Applicant	**0.028**	**-0.070**	**0.013**	**0.015**	**0.18*** **	**0.079*** **	**0.057**	**-0.027**	**0.0037***	**0.013*** **	**-0.0016**
After 2011*IDEX Awarded	**-0.0095**	**0.16**	**-0.00093**	**-0.023*** **	**-0.043**	**0.062*** **	**0.16*** **	**0.70*** **	**-0.00091**	**0.0037**	**0.00045**
Res. Seniority	0.12***	0.42***	0.095***	0.0047	0.064*	0.041***	0.13***	0.33***	0.0017	0.024***	0.0063***
Res. Seniority ^2	-0.0048***	-0.010***	-0.0028***	-0.00073***	-0.0042***	-0.00064***	-0.0013***	-0.0054***	-0.000063***	-0.00074***	-0.000094***
Constant	1.45***	2.86***	0.58***	0.34***	2.41***	-0.013	-0.39**	0.88	0.0081	0.0068	-0.033**
Observations	181,740	181,740	181,740	181,740	181,740	181,740	181,740	181,740	181,740	181,740	181,740
R-squared	0.010	0.002	0.011	0.009	0.010	0.003	0.007	0.003	0.000	0.009	0.001
Number of researchers	18,174	18,174	18,174	18,174	18,174	18,174	18,174	18,174	18,174	18,174	18,174
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the whole sample of active French researchers, including 329,470 researcher-year pairs, 32,947 researchers observed for ten years each, divided into three research fields. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variables *IDEX Applicant, IDEX Awarded,* and *After 2011*. However, these dummies are collinear to the researcher and year fixed effects and are not included in the regressions.

Table 11. Indirect effect of collaborating with IDEX awarded for non-applicant researchers, by field. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
HEALTH SCIENCES	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*Connect. to Awarded	**0.13**	**0.43**	**0.037**	**-0.017**	**0.38**	**-0.032**	**0.75*** **	**1.13**	**0.0087**	**-0.0012**	**0.00057**
Connected to awarded	0.0015	-0.81	-0.028	0.0061	0.18	-0.025	-0.52***	-0.42	0.0054	0.0031	0.0033
Res. Seniority	0.25*	0.89	0.21***	0.097**	-0.43	0.0034	0.32	0.33	0.0065	0.046***	-0.0032
Res. Seniority ^2	-0.0014	-0.00069	-0.00037	0.000026	-0.000019	0.00092	0.0044**	0.0039	-0.000048*	-0.00025***	0.000017
Constant	-0.49	-4.28	-1.58**	-0.65	8.95**	0.24	-3.59	-2.87	-0.062	-0.44***	0.038
Observations	6,500	6,500	6,500	6,500	6,500	6,500	6,500	6,500	6,500	6,500	6,500
R-squared	0.006	0.006	0.006	0.010	0.010	0.003	0.038	0.012	0.004	0.005	0.001
Number of researchers	650	650	650	650	650	650	650	650	650	650	650
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LIFE SCIENCES											
After 2011*Connect. to Awarded	**-0.0061**	**-0.95*** *	**-0.086**	**0.065***	**0.33**	**-0.021**	**0.43*** **	**0.30**	**0.00048**	**-0.011**	**-0.0035**
Connected to awarded	0.061	0.35	0.062	-0.042	0.017	0.087*	-0.17*	0.11	0.0011	0.013	-0.0015
Res. Seniority	0.028	0.094	0.050*	-0.025	-0.075	0.021	0.27***	0.20*	0.0038	0.027***	0.0051***
Res. Seniority ^2	-0.0029***	-0.0061**	-0.0013**	-0.00050**	-0.0022	0.000022	-0.0014	-0.0023	-0.000020	-0.00056***	-0.00011***
Constant	1.79***	4.91**	0.52**	0.80***	4.40***	-0.10	-2.08***	-0.41	-0.029	-0.12**	-0.024
Observations	9,080	9,080	9,080	9,080	9,080	9,080	9,080	9,080	9,080	9,080	9,080
R-squared	0.008	0.007	0.009	0.007	0.025	0.005	0.026	0.010	0.002	0.009	0.003
Number of researchers	908	908	908	908	908	908	908	908	908	908	908
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

PHYSICAL SCIENCES	(1) Pubs	(2) Citation Weighted Pubs	(3) Interdisciplinary Pubs	(4) Interdisciplinary Co-authors	(5) Within Lab Co-authors	(6) Within University Co-authors	(7) National Co-authors	(8) International Co-authors	(9) At least one Patent	(10) At least one Ph.D. Student	(11) At least one ANR Grant
After 2011*Connect. to Awarded	**-0.048**	**-0.62**	**0.022**	**-0.011**	**0.42*** **	**-0.082*** *	**0.11**	**1.22*** **	**0.0049**	**-0.00013**	**0.0040**
Connected to awarded	0.099	0.44	0.059	0.0035	-0.13	0.052	0.061	-0.0069	-0.00051	0.0064	-0.0054*
Res. Seniority	0.13***	0.35**	0.14***	0.021*	0.18***	-0.014	0.082*	0.55**	0.0043***	0.039***	0.00097
Res. Seniority ^2	-0.0045***	-0.0067	-0.0035***	-0.00068***	-0.0053***	-0.000044	-0.00083	-0.010**	-0.00011***	-0.00067***	-0.000049**
Constant	0.96***	1.07	0.0060	0.15	1.49***	0.27*	-0.044	-1.90	-0.011	-0.17	0.010
Observations	19,760	19,760	19,760	19,760	19,760	19,760	19,760	19,760	19,760	19,760	19,760
R-squared	0.009	0.002	0.015	0.009	0.008	0.003	0.003	0.004	0.002	0.013	0.001
Number of researchers	1,976	1,976	1,976	1,976	1,976	1,976	1,976	1,976	1,976	1,976	1,976
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample composed of 35,340 non-applicant researcher-year pairs, 3,534 researchers observed for ten years each, divided into three research fields. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variable *After 2011*. However, this dummy is collinear to the year fixed effects and is not included in the regressions.

2.8 Robustness checks

2.8.1 Mitigating the endogeneity problem

The estimation of the IDEX impact using the models of Equations 2 and 3 may be biased if there are time-variant factors that affect the researchers' outcomes of our treated and control groups differently and, at the same time, might relate to the IDEX endowment. For instance, universities may hire high-quality researchers during our study period, or obtain new resources from grants different than IDEX, such as ANR grants. The change in the environment where researchers work affects their performance and the characteristics of awarded universities might change differently than non-awarded universities (Lepori et al., 2019; McHale et al., 2022).

For this reason, we propose a robustness check where we include a set of time-variant university characteristics in the regressions of Equations 2 and 3. Specifically, we include the variable *University Size* that counts in year *t* the total number of active researchers affiliated to the university, the variable *University Affiliate Citation Weighted Publications* that measures in year *t* the average number of citation-weighted publications of the active university affiliates during the four preceding years, and the variable *University ANR Grants* that calculates in year *t* the total number of ANR grants obtained during the four preceding years by the active university affiliates[62]. When calculating these three variables, we exclude the focal researcher that is the unit of analysis. In Appendix G, we report the equations of the new models for the IDEX direct and indirect effects (Equations G1 and G2), and the respective OLS estimates (Tables G1 and G2). The results of the new estimations are in line with those of the main analyses presented in Table 8 and Table 9, with only two exceptions for the direct effect of IDEX. The coefficient of the effect of being awarded IDEX relative to applicant-non-award researchers loses its statistical significance in the regression explaining the number of co-authors within the same lab. In contrast, the coefficient of the effect of applying for IDEX relative to non-applicant researchers becomes significant in the regression explaining the number of national co-authors. The results on the indirect effect of IDEX are unchanged.

Another possible estimation issue concerns comparing groups of researchers in applicant, awarded, and non-applicant universities having different academic status and productivity trends. To mitigate this problem in identifying the IDEX effect, we propose a robustness check where we select a matched control for each researcher awarded IDEX. Specifically, for each researcher awarded IDEX, we consider a comparable researcher from a university that did apply but was not awarded IDEX. With this aim, we rely on the *Coarsened Exact Matching* (CEM) procedure (Iacus et al., 2012;

[62] We use the same definition of active researcher as in the main analysis. A researcher is active in year *t* if she has at least one publication in the previous four years, i.e., from *t-4* to *t-1*.

Azoulay, 2010). CEM appears more appropriate than the more common Propensity Score Matching for our framework because it is a nonparametric procedure. As argued in the methodology section 2.5, it is difficult to rely on a set of researchers' covariates to predict the probability that a university is awarded or not IDEX funding as the Propensity Score Matching does. CEM procedure matches researchers on pre-IDEX values for their stock of citation weighted publications, stock of the number of distinct co-authors, stock of ANR grants received, trend in citation weighted publications, trend in the number of distinct co-authors, trend in ANR grants received, seniority, field of research, and productivity breaks[63]. We coarsen the support of the joint distribution of the matching variables into 5,969 strata and allocate each researcher to a unique stratum. Researchers awarded IDEX are matched with applicant-non-award researchers allocated in the same stratum. This procedure guarantees the balance of the matching covariates between the IDEX awarded researchers and the applicant-non-award researchers. Table H1 of Appendix H reports the balance table pre- and post-matching. The CEM matching implies a notable reduction in our study sample. We remain with 5,969 awarded researchers matched with 5,969 similar applicant-non-awarded researchers, observed for ten years each. Moreover, this approach limits us to estimate only the effect of being awarded IDEX relative to applicant-non-award researchers. The advantage lies in the greater identification power of the estimation approach since we use now a *Conditional* (on CEM matching) *Difference-in-differences*. In Appendix H, we report the equation of the model estimated using the CEM sample in a difference-in-differences framework (Equation H1). Table H2 of Appendix H shows the OLS estimates of the effect of IDEX using the CEM sample. The results are in line with those reported in the main analysis of Table 8 concerning the IDEX direct effect relative to the group of IDEX applicant-non-awarded researchers (second row). There are only two exceptions. The coefficient of the effect of being awarded IDEX relative to applicant-non-award researchers loses its statistical significance in the regressions explaining the number of interdisciplinary publications and the number of co-authors within the same lab.

2.8.2 Disentangling the indirect effect of IDEX

When measuring the indirect effect of IDEX (Table 9), we consider the possible spillovers toward non-applicant researchers when connected with awarded researchers. We claimed that spillovers in scientific collaboration networks due to funding occur when non-award researchers are connected to researchers who obtain the funds. In this robustness check, we propose an exercise where we

[63] The stock variables are calculated by cumulating values from 1990 to 2010, one year before IDEX. The trend variables are calculated as the variation during the 5 years preceding IDEX, i.e., from 2006 to 2010. Productivity breaks count the number of years without any publication from 2006 to 2010 (Mairesse and Pezzoni, 2015). See Appendix H for further details.

investigate possible indirect effects due to the connection with researchers from universities that applied for IDEX funding but did not obtain it. Specifically, we disentangle the effect of being connected with an awarded researcher from the effect of being connected with an applicant researcher. The connection to applicants is measured by the new variable *Connected to Applicant*, a dummy that equals one in year *t* if the non-applicant researcher has a publication co-authored with an applicant researcher in the four preceding years, zero otherwise. The indirect effect of being connected to an applicant researcher is estimated by the coefficient of the interacted variable *After 2011*Connected to Applicant*, as described in Equation I1 of Appendix I. The additional effect of being connected to an awarded researcher is measured by the coefficient of the interacted variable *After 2011*Connected to Awarded*. The sample remains that of Table 9, composed of 3,534 non-applicant researchers. Table I1 of Appendix I reports the OLS estimates of the model of Equation I1. Interestingly, we find some indirect effects also when non-applicant researchers are connected to applicant researchers, regardless of the result of the application. Specifically, we find that the positive effect on the international network appears during the IDEX application (+0.45 international co-authors per year for non-applicant researchers if connected to applicant researchers, relative to researchers non-applicant and non-connected). Moreover, we find a negative effect on the number of publications of non-applicant researchers when connected to IDEX applicant researchers (-0.14 publications per year), and a positive effect on the number of interdisciplinary co-authors (+0.038 co-authors per year).

2.8.3 Removing the most productive researchers

In this robustness check, we re-estimate the models of Equations 2 and 3, explaining the direct and indirect effects of IDEX, removing the most productive researchers from our sample. In doing so, we want to explore whether the most productive researchers drive the effects of IDEX. We exclude from our sample researchers with productivity equal to or greater than 50 publications over our ten-year time window. In doing so, we remove 10.15% of researchers. We remain with a sample of 29,602 researchers observed for ten years each, divided into 26,278 applicant researchers and 3,324 non-applicant researchers, and 17,413 awarded researchers. Table J1 of Appendix J reports the OLS estimates of the model described in Equation 2, for the IDEX direct effect using the new restricted sample. Results are in line with those of Table 8, with two interesting exceptions. When we remove the most productive researchers, the coefficient of the effect of applying for IDEX loses its statistical significance and the coefficient of the effect of being awarded IDEX turns negative in the regression explaining the number of co-authors within the same laboratory. Moreover, the coefficient of the effect of being awarded IDEX turns significant and slightly negative in the regression explaining the number of publications. This means that the direct effect of IDEX on the 90% of researchers that are not among the most productive ones remains beneficial to enlarge the network outside the lab (within university, national, and international), but has the cost of decreasing the number of publications and the number of co-authors within the same lab. These two latter outcomes are driven by the 10% of most productive researchers. Also, the results of the indirect effect of IDEX estimated on the restricted sample (Table J2) are in line with those of Table 9, with two exceptions: non-applicant researchers that are connected with awarded researchers do not increase their national network (outside their university) and decrease their citation-weighted publication productivity when removing the 10% of most productive researchers.

2.8.4 Poisson pseudo-maximum likelihood estimator

Finally, we estimate the models described in Equations 2 and 3 using the Poisson pseudo-maximum likelihood estimator (PPML) with high-dimensional fixed effects, as described in Correia et al. (2020).

PPML presents some advantages over the OLS relative to our data. Specifically, it is well suited for counting data with discrete and non-negative dependent variables and a large number of zeroes. Moreover, Correia et al. (2020) show that PPML is appropriate for multiple fixed effects and interaction terms. However, the interpretation of the PPML estimates in a difference-in-differences framework is not straightforward since they have to be interpreted as the ratio of ratios rather than a difference in differences. Thus, we preferred to present OLS estimates in our main analysis.

In Appendix K, we report the equations of the models used to estimate the IDEX direct (Equation K1) and indirect (Equation K2) effects relying on the PPML estimator. As dependent variables, we consider only the researcher outcomes measured by discrete variables[64]. Table K1 and Table K2 of Appendix K report the PPML estimates. The results are in line with those of the main analyses presented in Table 8 and Table 9, with only a few exceptions. In particular, when estimating the direct effect of IDEX with PPML, the coefficient of the effect of being awarded IDEX loses its statistical significance in the regression explaining the number of national co-authors per year (outside the university). Concerning the indirect effect of IDEX, the effect of being connected with awarded researchers for non-applicant researchers remains positive and significant only in the regression explaining the number of international co-authors.

[64] For binary outcomes, we could have opted for other estimators, such as logistic regressions. However, the coefficients of the interaction terms in a logistic regression are not readily interpretable in a difference-in-differences framework (Blundell and Dias, 2009; Karaca-Mandic et al., 2012).

2.9 Conclusion

The change of university funding towards a competitive mechanism with quasi-market incentives has increased the pressure to document the effectiveness of government spending. The governments use funds to control the universities' research, education, and technology transfer agendas to align them to modern societal challenges and foster the country's international competitiveness (Geuna, 2001; Geuna and Rossi, 2015). However, literature aiming to quantify the impact of government university funding is still scant and far from reaching a consensus on what extent government funding might affect universities' outcomes.

This paper aims to fill this gap by evaluating the IDEX *Initiative d'excellence* program launched by the French government in 2011, to reward a group of selected universities with the necessary resources to become international centers of excellence.

We analyze the impact of IDEX on a broad set of outcomes of researchers affiliated with applicant and awarded universities. Moreover, we consider the indirect effect on researchers in universities that did not apply for IDEX but are connected with awarded researchers. We investigate eleven aspects of French researchers' activities that IDEX is expected to influence: quantity, quality, and interdisciplinarity of the publication productivity, interdisciplinarity in the collaborative behavior, scientific collaborations within the laboratory, scientific collaborations within the university, national scientific collaborations, international scientific collaborations, patenting, mentoring Ph.D. students, and fundraising.

We find that applying and being awarded IDEX affect researchers' outcomes. Researchers affiliated with universities applying for IDEX increase the number of interdisciplinary co-authors (+0.019 co-authors per year), the number of co-authors within the same laboratory (+0.14), and the number of co-authors within the same university but outside the lab (+0.10), relative to researchers in universities that did not apply. If the university is awarded IDEX, they further increase the number of interdisciplinary publications (+0.041 articles per year), the number of co-authors within the lab (+0.088), the number of co-authors within the same university but outside the lab (+0.090), the number of national co-authors outside their university (+0.35), and the number of international co-authors (+1.06), relative to researchers in applicant universities that did not obtain IDEX funding. When awarded, they decrease the number of interdisciplinary co-authors (-0.020). Furthermore, we find positive indirect effects of IDEX on researchers from universities that did not apply. If they are connected with awarded researchers, they increase the number of co-authors within the same lab (+0.42), the number of national co-authors outside their university (+0.30), and the number of international co-authors (+0.94), relative to non-applicant researchers who are not connected to researchers awarded IDEX.

Considering the magnitude of the coefficients, we find a relevant effect on the internationalization of the researchers' network when universities obtain IDEX funding: both awarded researchers and non-applicant researchers connected to awarded researchers gain around one new additional international co-author per year, relative to researchers in, or connected to, universities non awarded IDEX.

We find heterogeneity of our results across research fields but with consistent evidence of the impact of IDEX on researchers' networks. IDEX appears to be particularly beneficial for researchers in Health Sciences. The award of IDEX provides an important additional boost for their national collaborations outside their university (+0.85 co-authors per year) and international collaborations (+2.27). For Life Sciences and Physical Sciences, the main benefits occur for researchers' networks, and the highest impact concerns the increased size of the international network. When looking at the indirect effect of IDEX by field of research, the researchers' network is again the most influenced. Noteworthy, researchers from non-applicant universities in Physical Sciences experience an increase of 1.22 international co-authors per year if connected to awarded researchers, relative to non-connected researchers from non-applicant universities.

The fact that we found a positive effect of IDEX only on the size of the researchers' network might have different explanations. First, IDEX is a complex and ambitious program that aims to influence many research activities and implies a change in universities' structure and roles. These external constraints are likely to create conflicts within universities and increase the time universities need to adapt to the new funding mechanism and be effective in their new role (Geuna, 2001; De Boer et al., 2017). This scenario might explain the non-significant effects on researchers' publication productivity, Ph.D. mentoring, fundraising, and patenting activities. Consistent with our findings, Lanoë (2018) reported no effects of IDEX on researchers' publication impact, novelty, and diffusion, when analyzing the case of the University of Bordeaux. Second, the internationalization of French universities is one of the main goals of IDEX. This process can take several years. Collaborations are likely to be the first visible output, while publication productivity and other research benefits might emerge only after international collaborations are established. The creation of new international collaborations is still a good sign as, in modern science, high-impact research is increasingly the result of a collaborative effort (Wuthcy et al., 2007). Third, activities such as interdisciplinary research, entry into new research fields, and technology transfer are hardly stimulated by public funding even when they are not conditioned to a change in the university structure (Bruce et al., 2004; Hottenrott and Lawson, 2017; Kelchtermans et al., 2021; Carpentier, 2022). Finally, it is still unclear whether the concentration of resources to exploit economies of scale is a mechanism that works well within universities. The effectiveness of funding for science has shown to decrease due to obstacles that are scale-related, and research communities might operate better when they are small (Stephan, 2012).

This work has some limitations. First, we do not investigate the outcomes in the long-term due to data constraints and to not overlap with the second round of IDEX that is currently underway. Second, even if we have tried to proxy all the possible researchers' outcomes related to IDEX, we do not investigate other types of relationships with non-academic organizations, namely researchers' academic engagement (Perkmann et al., 2013). Finally, since we are measuring an average effect on the French university system, we do not explore heterogeneity in the IDEX treatment effect across universities (Cengiz et al., 2019). As follow-up work, we intend to explore the dynamics within universities by identifying the researchers who benefited most from IDEX and investigate whether their behaviors may explain some of the results we found.

2.10 Appendix of Chapter 2

APPENDIX G

This appendix reports the equations of the models, and their OLS estimates, used to assess the direct and indirect effect of IDEX for the robustness check where we include a set of time-variant university characteristics among the regressors. Specifically, we include the variables *University Size*, *University Affiliate Citation Weighted Publications*, and *University ANR Grants*, that proxy for the universities' size, quality, and fundraising.

Equation G1 reports the model we use to evaluate the direct effect of applying and being awarded IDEX, estimated on the sample of 329,470 researcher-year pairs. The effect of applying for IDEX is estimated by the coefficient of the interacted term *After 2011*IDEX Applicant*, while the effect of being awarded IDEX is estimated by the coefficient of the interacted term *After 2011*IDEX Awarded*. The model includes, on the right-hand side, the researcher's seniority, its squared term, and the three new variables describing the universities' characteristics. As the main model of Equation 2, we include year fixed effects (γ_t) and researcher fixed effects (α_i). $\varepsilon_{i,t}$ is the idiosyncratic error term[65]. Equation G2 details the model we use to assess the indirect effect of being connected with researchers awarded IDEX for non-applicant researchers, estimated on the sample of 35,340 non-applicant researcher-year pairs. The indirect effect of IDEX is estimated by the coefficient of the interacted term *After 2011*Connected to awarded*. The model includes, on the right-hand side, the researcher's seniority, its squared term, and the three new variables describing the universities' characteristics. As the main model of Equation 3, we include year fixed effects (γ_t) and researcher fixed effects (α_i). $\varepsilon_{i,t}$ is the idiosyncratic error term[66].

OLS estimates of the models described in Equation G1 and Equation G2 are reported in Table G1 and Table G2.

$$
\begin{aligned}
&Researcher\ outcome_{i,t} \\
&\quad = \beta_0 + \beta_1\ After\ 2011 * IDEX\ Applicant_i + \beta_2\ After\ 2011 * IDEX\ Awarded_i \\
&\quad + \beta_3\ Researcher's\ Seniority_{i,t} + \beta_4\ Researcher's\ seniority^2{}_{i,t} \\
&\quad + \beta_5\ University\ Size_{i,t} \\
&\quad + \beta_6\ University\ Affiliate\ Citation\ Weighted\ Publications_{i,t} \\
&\quad + \beta_7\ University\ ANR\ Grants_{i,t} + \gamma_t + \alpha_i + \varepsilon_{i,t}
\end{aligned}
$$

Equation G1

[65] A difference-in-differences exercise requires to include in the equation the non-interacted dummy variables *IDEX Applicant*, *IDEX Awarded*, and *After 2011*. However, these dummies are collinear to the researcher and year fixed effects and are not included in the model of Equation G1.

[66] Equation G2 reports the non-interacted variable *Connected to awarded* since, being this variable time-variant, it is not collinear to the researcher fixed effects. As for Equation G1, the non-interacted variable *After 2011* is collinear to the year fixed effects and is not included in the regression of Equation G2.

$$
\begin{aligned}
Researcher\ outcome_{i,t} &= \beta_0 + \beta_1\ After\ 2011 * Connected\ to\ awarded_{i,t} \\
&+ \beta_2\ Connected\ to\ awarded_{i,t} + \beta_3\ Researcher's\ Seniority_{i,t} \\
&+ \beta_4\ Researcher's\ seniority^2{}_{i,t} + \beta_5\ University\ Size_{i,t} \\
&+ \beta_6\ University\ Affiliate\ Citation\ Weighted\ Publications_{i,t} \\
&+ \beta_7\ University\ ANR\ Grants_{i,t} + \gamma_t + \alpha_i + \varepsilon_{i,t}
\end{aligned}
$$

Equation G2

Table G1. Direct effect of applying and being awarded IDEX on French researchers' outcomes, including university controls. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*IDEX Applicant	**0.039**	**-0.17**	**0.015**	**0.031*** **	**0.25*** **	**0.099*** **	**-0.16*** **	**-0.22**	**0.00085**	**0.0067* **	**-0.0015**
After 2011*IDEX Awarded	**0.026**	**0.079**	**0.030* **	**-0.013** **	**0.070**	**0.070*** **	**0.30*** **	**0.95*** **	**-0.00052**	**-0.0040**	**-0.0014**
Res. Seniority	0.11***	0.50***	0.081***	0.0029	0.093***	0.052***	0.16***	0.27***	0.00092	0.020***	0.0063***
Res. Seniority ^2	-0.0044***	-0.012***	-0.0023***	-0.00073***	-0.0030***	-0.00043***	0.00020	-0.00076	-0.000044***	-0.00063***	-0.00010***
University Size	6.0e-07	-0.000021	6.2e-07	-4.3e-06**	-8.6e-06	-7.9e-06	0.000026**	0.000087**	-9.1e-08	1.4e-06*	4.9e-07
Univ. affiliate citation-weighted pubs	0.013	0.063	0.023	0.0039	0.19***	0.040**	-0.041	-0.26*	-0.0021	0.017***	0.0011
University ANR grants	0.000043	0.0036***	0.000027	-0.000045*	-0.0011***	0.00044***	0.00082***	0.0023***	0.000012*	-0.00004***	-0.000024***
Constant	1.51***	3.24***	0.47***	0.50***	2.11***	-0.23**	-1.18***	0.0079	0.020	-0.057	-0.039**
Observations	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470
R-squared	0.009	0.002	0.010	0.008	0.012	0.005	0.016	0.007	0.000	0.008	0.001
Number of researchers	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the whole sample of active French researchers, including 329,470 researcher-year pairs, 32,947 researchers observed for ten years each. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variables *IDEX Applicant, IDEX Awarded*, and *After 2011*. However, these dummies are collinear to the researcher and year fixed effects and are not included in the regressions.

Table G2. Indirect effect of collaborating with IDEX awarded for non-applicant researchers, including university controls. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*Connect. to Awarded	**-0.0018**	**-0.53**	**-0.00063**	**0.010**	**0.42*** **	**-0.052**	**0.30*** **	**0.94*** **	**0.0048**	**-0.0031**	**0.0013**
Connected to Awarded	0.069	0.17	0.041	-0.0087	-0.046	0.047*	-0.12	-0.062	0.00074	0.0072	-0.0027
Res. Seniority	0.083***	0.28**	0.087***	0.0091	0.16**	-0.0076	0.11***	0.35***	0.0038***	0.035***	0.0014
Res. Seniority ^2	-0.0034***	-0.0053**	-0.0023***	-0.00048***	-0.0031***	0.00020	0.00051	-0.0049**	-0.000069***	-0.00053***	-0.000050***
University Size	0.000078	0.00012	-3.4e-06	0.000038**	0.00045***	0.000059**	-0.000025	0.000023	-3.4e-06	-4.5e-06	2.3e-07
Univ. affiliate citation-weighted pubs	0.089	0.12	-0.011	0.0099	0.16	-0.026	0.100	-0.087	0.0043	0.016**	0.00047
University ANR grants	-0.00053	-0.0030	0.00072	-0.00026	-0.0037	-0.00021	0.0024	-0.0021	0.000091	-0.000013	-0.000052
Constant	0.92***	1.54	0.35	0.18	-0.011	0.10	-0.68	-0.91	-0.015	-0.21	0.0030
Observations	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340
R-squared	0.007	0.002	0.010	0.007	0.010	0.001	0.010	0.004	0.001	0.010	0.001
Number of researchers	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample composed of 35,340 non-applicant researcher-year pairs, 3,534 researchers observed for ten years each. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variable *After 2011*. However, this dummy is collinear to the year fixed effects and is not included in the regressions.

APPENDIX H

This appendix reports the description and the balance table of the researcher variables used for the *Coarsened Exact Matching* (CEM) for the robustness check where we estimate the IDEX direct effect on a sample of awarded and similar applicant-not-award researchers. Then, we report the model of the *Conditional* (on CEM matching) *Difference-in-differences* and its respective OLS estimates.

The CEM matching aims to pair researchers awarded IDEX with similar researchers from universities that did apply but were not awarded IDEX, based on a set of researcher variables. These variables describe the researchers' stock and trends in productivity before IDEX. We include the *Stock of Citation Weighted Publications* that measures the cumulative number of the researcher's quality-weighted publications from 1990 to 2010; the *Stock of Co-authors* that measures the cumulative number of the researcher's total distinct co-authors from 1990 to 2010; the *Stock of ANR Grants* that measures the cumulative number of ANR grants received by the researcher from 2005[67] to 2010; the *Trend in Citation Weighted Publications* that calculates the variation in the cumulative number of researcher's quality-weighted publications between 2006 and 2010; the *Trend in Co-Authors* that calculates the variation in the cumulative number of the researcher's total distinct co-authors between 2006 and 2010; the *Trend in ANR Grants* that calculates the variation in the cumulative number of ANR grants received by the researcher between 2006 and 2010; the researcher's *Seniority* that counts the years elapsed between the researcher's first publication and 2010; the researcher's *Productivity Breaks* that count the number of years between 2006 and 2010 without any researcher's publication to account for the continuity in her productivity (Mairesse and Pezzoni, 2015); and three dummies indicating the field of research to which a researcher belongs, namely *Health Sciences*, *Life Sciences,* and *Physical Sciences*.

Table H1 reports the balance table of the researcher variables pre- and post-CEM matching. After the matching, we end up with 5,696 awarded researchers and 5,696 similar applicant-non-award researchers. All the researcher variables are statistically equivalent between awarded and applicant-non-award researchers after the matching.

Table H1. Means of the pre-IDEX variables for awarded researchers and applicant-non-award researchers, before (Columns 1 and 2) and after (Columns 3 and 4) the CEM.

	Sample before CEM			CEM sample		
	(1) **Awarded**	(2) **Applicant non-awarded**	**p-value**	(3) **Awarded**	(4) **Applicant non-awarded**	**p-value**
N. of researchers	19,753	9,660		5,969	5,969	
Researcher variables						
Stock of Citation Weighted Publications	75.76***	59.09	0.000	34.70	34.20	0.468
Stock of Co-authors	82.39***	62.36	0.000	40.14	40.26	0.877
Stock of ANR Grants	0.118***	0.103	0.001	0.026	0.026	1.000
Trend in Citation Weighted Publications	25.55***	20.96	0.000	11.87	11.78	0.719
Trend in Co-Authors	26.02***	19.86	0.000	12.26	12.32	0.802
Trend in ANR Grants	0.079	0.073	0.113	0.020	0.020	1.000
Seniority	12.91***	12.70	0.002	12.19	12.19	0.976
Productivity Breaks	1.29***	1.35	0.000	1.68	1.68	1.000
Health Sciences	0.17***	0.11	0.000	0.07	0.07	1.000
Life Sciences	0.31***	0.26	0.000	0.24	0.24	1.000
Physical Sciences	0.52***	0.63	0.000	0.69	0.69	1.000

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

[67] The stock of ANR grants is calculated starting from 2005 since the ANR agency was created in that year.

Equation H1 describes the model of the *Conditional Difference-in-differences* that we use to assess the direct effect of being awarded IDEX, estimated on the CEM sample of 119,380 researcher-year pairs. This sample is composed of 11,938 researchers observed for ten years each and divided into 5,969 awarded researchers and 5,969 applicant but not awarded similar researchers. The effect of being awarded IDEX, relative to applicant-non-award researchers, is estimated by the coefficient of the interacted term *After 2011*IDEX Awarded*. As the main model of Equation 2, we include year fixed effects (γ_t) and researcher fixed effects (α_i). $\varepsilon_{i,t}$ is the idiosyncratic error term[68].

OLS estimates of the models described in Equation H1 are reported in Table H2.

$$Researcher\ outcome_{i,t} = \beta_0 + \beta_1\ After\ 2011 * IDEX\ Awarded_i + \gamma_t + \alpha_i + \varepsilon_{i,t}$$

Equation H1

[68] A difference-in-differences exercise requires to include in the equation the non-interacted dummy variables *IDEX Awarded* and *After 2011*. However, these dummies are collinear to the researcher and year fixed effects and are not included in the model of Equation H1. Differently from Equation 2, Equation H1 does not include the researcher's seniority and its squared term, being the variable *Seniority* among the variables used for the Coarsened Exact Matching.

Table H2. Direct effect of being awarded IDEX on French researchers' outcomes, using the CEM sample. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*IDEX Awarded	**-0.016**	**0.11**	**0.016**	**-0.018****	**-0.018**	**0.059*****	**0.089*****	**0.23*****	**-0.00080**	**-0.000034**	**0.0014**
Constant	1.38***	2.96***	0.77***	0.30***	1.98***	0.20***	0.42***	1.07***	0.010***	0.11***	0.0041***
Observations	119,380	119,380	119,380	119,380	119,380	119,380	119,380	119,380	119,380	119,380	119,380
R-squared	0.003	0.003	0.005	0.006	0.009	0.004	0.011	0.008	0.000	0.002	0.001
Number of researchers	11,938	11,938	11,938	11,938	11,938	11,938	11,938	11,938	11,938	11,938	11,938
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample of researchers obtained using the CEM matching. It includes 119,380 researcher-year pairs, 11,938 researchers observed for ten years each, divided into 5,969 awarded researchers and 5,969 applicant but not awarded similar researchers. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variables *IDEX Awarded* and *After 2011*. However, these dummies are collinear to the researcher and year fixed effects and are not included in the regressions.

APPENDIX I

This appendix reports the equation of the model used for the robustness check where we aim to disentangle the indirect effect of IDEX, estimated on the sample of 35,340 non-applicant researcher-year pairs, and its OLS estimates. Specifically, we include in the model of Equation I1 the variable *Connected to Applicant.* We estimate the effect of being connected with IDEX applicants on non-applicant researchers by the coefficient of the interacted term *After 2011*Connected to applicant*. We estimate the effect of being connected with IDEX awarded researchers by the coefficient of the interacted term *After 2011*Connected to awarded.* The model includes, on the right-hand side, the researcher's seniority and its squared term, and the year fixed effects (γ_t) and researcher fixed effects (α_i). $\varepsilon_{i,t}$ is the idiosyncratic error term[69].

OLS estimates of the model described in Equation I1 are reported in Table I1.

$$\begin{aligned}
Researcher\ outcome_{i,t} \\
&= \beta_0 \\
&+ \beta_1\ After\ 2011 * Connected\ to\ applicant_{i,t} \\
&+ \beta_2\ After\ 2011 * Connected\ to\ awarded_{i,t} + \beta_3\ Connected\ to\ applicant_{i,t} \\
&+ \beta_4\ Connected\ to\ awarded_{i,t} + \beta_5\ Researcher's\ Senioriy_{i,t} \\
&+ \beta_6\ Researcher's\ Senioriy^2{}_{i,t} + \gamma_t + \alpha_i + \varepsilon_{i,t}
\end{aligned}$$

Equation I1

[69] A difference-in-differences exercise requires to include in the equation the non-interacted dummy variable *After 2011*. However, this dummy is collinear to the year fixed effects and is not included in the model of Equation I1. Equation I1 reports the non-interacted variables *Connected to applicant* and *Connected to awarded* since, being these variables time-variant, they are not collinear to the researcher fixed effects.

Table I1. Indirect effect of collaborating with IDEX applicants and awarded for non-applicant researchers. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*Connect. to Applicant	**-0.14***	**-0.36**	**-0.036**	**0.038***	**0.23**	**-0.023**	**0.030**	**0.45***	**-0.0034**	**-0.0067**	**0.00015**
After 2011*Connect. to Awarded	**0.12**	**-0.25**	**0.030**	**-0.022**	**0.23**	**-0.032**	**0.27****	**0.57***	**0.0076**	**0.0023**	**0.0014**
Connected to Applicant	0.061	-0.13	0.028	-0.036**	-0.12	-0.0074	-0.15**	-0.17	0.0044	0.0094	0.0022
Connected to Awarded	0.018	0.27	0.018	0.021	0.057	0.052	0.0050	0.083	-0.0028	-0.00040	-0.0045
Res. Seniority	0.081***	0.28**	0.087***	0.0074	0.14**	-0.0085	0.10***	0.35***	0.0037***	0.034***	0.0015
Res. Seniority ^2	-0.0034***	-0.0053**	-0.0022***	-0.00050***	-0.0032***	0.00020	0.00048	-0.0050**	-0.000067***	-0.00053***	-0.000050***
Constant	1.39***	2.20**	0.31*	0.35***	1.98***	0.24*	-0.43	-1.00	-0.016	-0.19	0.0037
Observations	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340	35,340
R-squared	0.007	0.002	0.010	0.007	0.009	0.001	0.010	0.004	0.001	0.009	0.001
Number of researchers	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534	3,534
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample composed of 35,340 non-applicant researcher-year pairs, 3,534 researchers observed for ten years each. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variable *After 2011*. However, this dummy is collinear to the year fixed effects and is not included in the regressions.

APPENDIX J

This appendix reports the OLS estimates of the models described in Equation 2 and Equation 3, estimated on the samples of 296,020 active researcher-year pairs and 33,240 non-applicant researcher-year pairs, respectively. Differently from the main analyses of Table 8 and Table 9, these exercises exclude the most productive researchers. Most productive researchers are defined as those having a total productivity over the ten years of our time window equal to or greater than 50 publications. We re-estimate the direct effect of applying and being awarded IDEX in Table J1, and the indirect effect of being connected with researchers awarded IDEX for non-applicant researchers in Table J2.

Table J1. Direct effect of applying and being awarded IDEX on French researchers' outcomes, removing the most productive researchers. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*IDEX Applicant	**0.018**	**-0.088**	**-0.013**	**0.012**	**0.087**	**0.079*** **	**-0.053**	**0.100**	**0.0013**	**0.0042**	**-0.0010**
After 2011*IDEX Awarded	**-0.046*** **	**-0.13**	**-0.0031**	**-0.021*** **	**-0.066** **	**0.045*** **	**0.14*** **	**0.29*** **	**-0.00057**	**0.0013**	**-0.00042**
Res. Seniority	0.058***	0.24***	0.047***	0.0023	0.076**	0.044***	0.14***	0.18***	0.00070	0.018***	0.0059***
Res. Seniority ^2	-0.0043***	-0.012***	-0.0023***	-0.00079***	-0.0048***	-0.00070***	-0.0014***	-0.0053***	-0.000044***	-0.00059***	-0.000089***
Constant	1.63***	4.04***	0.68***	0.43***	2.46***	-0.094	-0.62***	1.05*	0.012	0.011	-0.030**
Observations	296,020	296,020	296,020	296,020	296,020	296,020	296,020	296,020	296,020	296,020	296,020
R-squared	0.011	0.003	0.009	0.007	0.008	0.003	0.009	0.003	0.000	0.007	0.001
Number of researchers	29,602	29,602	29,602	29,602	29,602	29,602	29,602	29,602	29,602	29,602	29,602
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample of active researchers excluding the most productive ones. It consists of 296,020 researcher-year pairs, 29,602 researchers observed for ten years each. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variables *IDEX Applicant, IDEX Awarded,* and *After 2011*. However, these dummies are collinear to the researcher and year fixed effects and are not included in the regressions.

Table J2. Indirect effect of collaborating with IDEX awarded for non-applicant researchers, removing the most productive researchers. OLS estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
After 2011*Connect. to Awarded	**-0.012**	**-0.44***	**-0.0074**	**0.011**	**0.23***	**-0.018**	**0.11**	**0.60*** **	**0.0049**	**0.00057**	**0.0013**
Connected to awarded	0.036	0.068	0.011	-0.018	-0.064	0.016	-0.048	0.098	-0.00062	0.0056	-0.0036*
Res. Seniority	0.054**	0.18**	0.065***	0.0080	0.13**	-0.0089	0.11***	0.22**	0.0028***	0.033***	0.00064
Res. Seniority ^2	-0.0033***	-0.0057***	-0.0020***	-0.00053***	-0.0036***	0.00011	-0.00028	-0.0042**	-0.000060***	-0.00049***	-0.000036**
Constant	1.38***	2.53***	0.36**	0.32***	1.77***	0.24*	-0.47	-0.017	-0.0083	-0.18	0.0093
Observations	33,240	33,240	33,240	33,240	33,240	33,240	33,240	33,240	33,240	33,240	33,240
R-squared	0.008	0.002	0.009	0.006	0.007	0.001	0.007	0.003	0.001	0.008	0.001
Number of researchers	3,324	3,324	3,324	3,324	3,324	3,324	3,324	3,324	3,324	3,324	3,324
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample of non-applicant researchers excluding the most productive ones. It consists of 33,240 non-applicant researcher-year pairs, 3,324 researchers observed for ten years each. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variable *After 2011*. However, this dummy is collinear to the year fixed effects and is not included in the regressions.

APPENDIX K

This appendix reports the equations and the estimates of the models where we implement the *Poisson pseudo-maximum likelihood* estimator to assess the direct and indirect effects of IDEX funding. Differently from Equation 2 and Equation 3 of the main analysis, in Equation K1 and Equation K2 the regressors are inside an exponential function. Moreover, on the left-hand side of the equations, we consider as researcher i's outcome only the discrete variables described in section 2.4.1, i.e., *Publications, Citation Weighted Publications, Interdisciplinary Publications, Interdisciplinary Co-authors, Within Lab Co-authors, Within University Co-authors, National Co-authors,* and *International Co-authors*.

Equation K1 reports the model we use to evaluate the direct effect of applying and being awarded IDEX, estimated on the sample of 32,947 researchers. The effect of applying for IDEX is estimated by the coefficient of the interacted term *After 2011*IDEX Applicant*, while the effect of being awarded IDEX is estimated by the coefficient of the interacted term *After 2011*IDEX Awarded.* The model includes, on the right-hand side, the researcher's seniority, its squared term, and the year fixed effects (γ_t) and researcher fixed effects (α_i). $\varepsilon_{i,t}$ is the idiosyncratic error term[70]. Equation K2 details the model we use to assess the indirect effect of being connected with researchers awarded IDEX for non-applicant researchers, estimated on the sample of 3,534 non-applicant researchers. The indirect effect of IDEX is estimated by the coefficient of the interacted term *After 2011*Connected to awarded*. The model includes, on the right-hand side, the researcher's seniority, its squared term, and the year fixed effects (γ_t) and researcher fixed effects (α_i). $\varepsilon_{i,t}$ is the idiosyncratic error term[71].

Table K1 and Table K2 report the PPML estimates of the models described in Equations K1 and K2. The estimated coefficients are reported transformed to incidence-rate ratios, i.e., exp(β_i) rather than β_i.

$$\begin{aligned} Researcher\ outcome_{i,t} &= \exp\,[\,\beta_0 + \beta_1\, After\ 2011 * IDEX\ Applicant_i \\ &+ \beta_2\, After\ 2011 * IDEX\ Awarded_i + \beta_3\, Researcher's\ Senioriy_{i,t} \\ &+ \beta_4\, Researcher's\ Senioriy^2{}_{i,t} + \gamma_t + \alpha_i\,] + \varepsilon_{i,t} \end{aligned}$$

Equation K1

$$\begin{aligned} Researcher\ outcome_{i,t} &= \exp\,[\,\beta_0 \\ &+ \beta_1\, After\ 2011 * Connected\ to\ awarded_{i,t} + \beta_2\, Connected\ to\ awarded_{i,t} \\ &+ \beta_3\, Researcher's\ Seniority_{i,t} + \beta_4\, Researcher's\ Seniority^2{}_{i,t} + \gamma_t + \alpha_i\,] \\ &+ \varepsilon_{i,t} \end{aligned}$$

Equation K2

[70] A difference-in-differences exercise requires to include in the equation the non-interacted dummy variables *IDEX Applicant*, *IDEX Awarded*, and *After 2011*. However, these dummies are collinear to the researcher and year fixed effects and are not included in the model of Equation K1.

[71] Equation K2 reports the non-interacted variable *Connected to awarded* since, being this variable time-variant, it is not collinear to the researcher fixed effects. As for Equation K1, the non-interacted variable *After 2011* is collinear to the year fixed effects and is not included in the regression of Equation K2.

Table K1. Direct effect of applying and being awarded IDEX on French researchers' outcomes. PPML estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors
After 2011*IDEX Applicant	**1.014**	**0.994**	**0.998**	**1.028**	**1.058*** **	**1.285*** **	**1.022**	**1.035**
After 2011*IDEX Awarded	**1.007**	**1.013**	**1.027** **	**0.961*** **	**0.980* **	**1.056* **	**0.995**	**1.105*** **
Res. Seniority	1.098***	1.071***	1.113***	1.107***	1.080***	1.130***	1.174***	1.118***
Res. Seniority ^2	0.998***	0.998***	0.997***	0.998***	0.999***	0.998***	0.998***	0.998***
Constant	1.552***	9.457***	0.778***	0.275***	2.431***	0.375***	0.401***	3.913***
Observations	328,350	327,805	314,226	259,506	320,777	181,638	267,547	293,807
Pseudo R-squared	0.3611	0.5840	0.3261	0.1586	0.3790	0.3109	0.4102	0.6257
Number of researchers	32,947	32,885	31,506	26,011	32,176	18,187	26,801	29,439
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the whole sample of 32,947 active French researchers. Coefficients are reported in exponentiated form. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variables *IDEX Applicant*, *IDEX Awarded*, and *After 2011*. However, these dummies are collinear to the researcher and year fixed effects and are not included in the regressions.

Table K2. Indirect effect of collaborating with IDEX awarded for non-applicant researchers. PPML estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pubs	Citation Weighted Pubs	Interdisciplinary Pubs	Interdisciplinary Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors
After 2011*Connect. to Awarded	**0.983**	**0.901***	**0.961**	**0.975**	**1.045**	**0.812***	**0.851**	**1.221****
Connected to Awarded	1.035	1.029	1.042*	0.994	1.005	1.180*	1.053	1.033
Res. Seniority	1.084***	1.056***	1.115***	1.088***	1.064***	1.035	1.128***	1.122***
Res. Seniority ^2	0.998***	0.999**	0.997***	0.998***	0.999***	1.001	0.999	0.998***
Constant	1.358***	5.533***	0.652***	0.289***	2.630***	0.463***	0.444***	2.569***
Observations	35,210	35,101	33,274	27,433	34,556	14,198	27,811	30,086
Pseudo R-squared	0.2905	0.4750	0.2792	0.1498	0.3328	0.2311	0.3571	0.5445
Number of researchers	3,534	3,522	3,337	2,750	3,468	1,422	2,786	3,014
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample composed of 3,534 non-applicant researchers. Coefficients are reported in exponentiated form. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variable *After 2011*. However, this dummy is collinear to the year fixed effects and is not included in the regressions.

APPENDIX L

This appendix discusses the identification strategy of the econometric approach we use to assess the effect of IDEX by testing for the common trends assumption. The common trends assumption is the key identifying assumption on which the *Difference-in-differences* approach that we use in our analyses relies. It implies that researchers in universities treated with IDEX would have similar trends in the outcomes to researchers in universities not treated with IDEX in the absence of IDEX. To formally test for that, we propose a regression based test augmenting the base model of Equation 2 with the leads and lags, following Autor (2003) and Angrist and Pischke (2009). Specifically, we construct the leads and lags as interactions between the treatment indicators and the year dummies, allowing therefore for multiple treatments. In doing so, we test for the common trends assumption by looking at the set of interactions before and after 2011, i.e., when IDEX occurred. If the pre-IDEX interactions are not statistically significant, it means that the common trends assumption is satisfied since there are no statistical differences in the pre-treatment trends. The interactions after IDEX inform about the dynamics of the IDEX effect on researchers' outcomes. We also test the joint statistical significance of the interactions before and after IDEX as further validation of the common trends assumption. Not rejecting the null hypothesis of no significant variation among the interactions before IDEX while rejecting the null hypothesis after IDEX assures the fulfillment of the common trends assumption.

Equations L1 and L2 report the models with the leads and lags that we use to test for the common trends assumption. Equation L1 is estimated on the main sample of 329,470 researcher-year pairs. Being t the treatment year, i.e., 2011, when IDEX occurred, we include in the model five periods before IDEX and four periods after IDEX. Then, we include the interactions between each period and the treatment indicators *IDEX Applicant* and *IDEX Awarded*. As in the main model of Equation 2, we also include the researcher's seniority and its squared term among the regressors, and the researcher fixed effects (α_i). Finally, $\varepsilon_{i,t}$ is the idiosyncratic error[72]. Table L1 reports the OLS estimates of the model described in Equation L1. When estimating the effect of applying for IDEX, the common trends assumption is satisfied in the regression explaining the number of *Within Lab Co-authors,* i.e., the number of researcher's co-authors within the same laboratory. In this regression, we observe no statistical differences in the pre-IDEX trends (from Table L2, the F-value of the joint significance of the interaction terms pre-IDEX is equal to 1.37) while we find statistically significant differences in the post-IDEX trends (the F-value of the joint significance of the interaction terms post-IDEX equal to 3.63). This finding is in line with the results of Table 8 of the main text, showing that applying for IDEX has a positive effect on the researcher's collaborations within the same laboratory, regardless of the result of the application. The post-IDEX trends provide a sense of the dynamics of the effect of applying for IDEX. In particular, the positive effect on the number of co-authors within the laboratory takes place two years after IDEX and remains rather stable in the following years. Following the same interpretation of pre and post-IDEX trends, Table L1 suggests that the common trends assumption is not satisfied in the estimation of the effect of being awarded IDEX. For this reason, we propose the model of Equation

[72] A difference-in-differences exercise requires to include in the equation the non-interacted dummy variables *IDEX Applicant* and *IDEX Awarded*. However, these dummies are collinear to the researcher fixed effects and are not included in the model of Equation L1.

L2 that we estimate on the sample of 119,380 researcher-year pairs obtained after applying the *Coarsened Exact Matching* (CEM), as explained in the robustness check of section 2.8.1. The results of the robustness check using the CEM matching are in line with those of the main analysis of Table 8, but differently from the main analysis, the CEM sample is created conditionally on the matching of pre-IDEX trends in researchers' outcomes. Thus, we expect that the common trends assumption is satisfied using the CEM sample. We test this assumption using the model described in Equation L2, in which we include the period dummies, the leads and lags, and the researcher fixed effects (α_i) similar to the model of Equation L1[73]. This model only estimates the effect of being awarded IDEX since the sample is composed of 11,938 researchers divided into 5,969 awarded researchers and 5,969 applicant but not awarded similar researchers. Table L3 reports the OLS estimates of the model described in Equation L2. We find that the common trends assumption is satisfied in the regressions explaining the number of *National Co-authors* and the number of *International Co-authors*. In these regressions, we observe no statistical differences in the pre-IDEX trends (from Table L4, the F-value of the joint significance of the interaction terms pre-IDEX equal to 0.9 in the regression explaining the *National Co-authors* and 1.53 in the regression explaining the *International Co-authors*) while we find statistically significant differences in the post-IDEX trends (the F-value of the joint significance of the interaction terms post-IDEX equal to 2.86 in the regression explaining the *National Co-authors* and 2.66 in the regression explaining the *International Co-authors*). These findings are in line with the results of Table 8 of the main text, showing that researchers in universities awarded IDEX increases their national and international collaborations relative to researchers in universities that applied but were not awarded IDEX. The dynamics of these effects are expressed by the post-IDEX trends and suggest that the positive effect of being awarded IDEX on the number of national co-authors takes place every year starting from the IDEX award, except for the second year after IDEX, with a rather stable magnitude. On the contrary, the positive effect of IDEX on the number of international co-authors is concentrated in the IDEX award year and the year immediately following, with a magnitude that is more than double the effect on national co-authors.

$$
\begin{aligned}
&Researcher\ outcome_{i,t} \\
&\quad = \beta_0 + \sum_{j=-5}^{4} \gamma_j\, period_{t+j} + \sum_{j=-5}^{4} \beta_j\ period_{t+j} * IDEX\ applicant_i \\
&\quad + \sum_{j=-5}^{4} \delta_j\ period_{t+j} * IDEX\ awarded_i + \varphi_1\ Researcher's\ Senioriy_{i,t} \\
&\quad + \varphi_2\ Researcher's\ Senioriy^2{}_{i,t} + \alpha_i + \varepsilon_{i,t}
\end{aligned}
$$

Equation L1, estimated using the main sample

[73] A difference-in-differences exercise requires to include in the equation the non-interacted dummy variable *IDEX Awarded*. However, this dummy is collinear to the researcher fixed effects and is not included in the model of Equation L2. Differently from Equation L1, Equation L2 does not include the researcher's seniority and its squared term, being the variable *Seniority* among the variables used for the Coarsened Exact Matching.

$$Researcher\ outcome_{i,t} = \beta_0 + \sum_{j=-5}^{4} \gamma_j\, period_{t+j} + \sum_{j=-5}^{4} \beta_j\ period_{t+j} * IDEX\ awarded_i + \alpha_i + \varepsilon_{i,t}$$

Equation L2, estimated using the CEM sample

Table L1. Direct effect of applying and being awarded IDEX on French researchers' outcomes. OLS estimates including leads and lags.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdiscipl. Pubs	Interdiscipl. Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
(period t-5)*IDEX Applicant	-0.041	-0.16	-0.012	0.0021	-0.078	-0.048*	0.023	0.00080	-0.0073**	-0.00066	0.0032
(period t-4)*IDEX Applicant	-0.096**	-0.42	0.012	0.0051	-0.043	-0.10***	-0.14	-0.048	-0.0060*	-0.0063	0.0034
(period t-3)*IDEX Applicant	0.031	-0.13	0.016	-0.0059	0.11	0.023	-0.0097	-0.11	-0.0053*	0.0011	0.0020
(period t-2)*IDEX Applicant	0.030	-0.28	0.023	0.025*	0.053	-0.065**	0.034	-0.42**	-0.0056*	-0.0092	0.00093
(period t-1)*IDEX Applicant	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
(period t)*IDEX Applicant	-0.022	-0.19	-0.0046	0.019	0.016	0.087***	-0.11*	0.11	-0.0018	-0.0087	0.0019
(period t+1)*IDEX Applicant	0.0085	0.0067	0.020	0.023	-0.040	0.040	-0.11	-0.31	-0.0037	0.0060	0.0025
(period t+2)*IDEX Applicant	0.065	-0.19	0.033	0.025	0.28***	0.080***	0.036	0.055	-0.00097	-0.0032	-0.0023
(period t+3)*IDEX Applicant	0.032	-0.32	0.015	0.016	0.19*	0.049	-0.12	0.29	-0.0052	-0.00034	-0.00047
(period t+4)*IDEX Applicant	0.042	-0.36	0.044	0.039**	0.31***	0.068**	-0.062	-0.086	-0.0042	0.014*	-0.0026
(period t-5)*IDEX Awarded	-0.16***	-1.65***	-0.18***	0.014	-0.14**	-0.090***	-0.48***	-1.49***	0.0023	0.0078	0.0022
(period t-4)*IDEX Awarded	-0.093***	-1.03***	-0.17***	0.020**	-0.063	-0.080***	-0.35***	-1.31***	-0.00027	-0.0072	-0.0037
(period t-3)*IDEX Awarded	-0.17***	-1.01***	-0.16***	0.0012	-0.19***	-0.064***	-0.25***	-1.08***	0.0027	0.0017	-0.0027
(period t-2)*IDEX Awarded	-0.017	0.056	-0.065***	0.0044	0.021	0.0042	0.069	-0.11	0.0011	0.0045	-0.00055
(period t-1)*IDEX Awarded	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
(period t)*IDEX Awarded	-0.0089	-0.061	-0.039*	-0.012	0.085	0.051**	0.15***	0.13	0.00050	0.00057	-0.0025
(period t+1)*IDEX Awarded	0.0057	-0.60**	-0.025	0.00016	0.26***	0.030	0.18***	0.46***	-0.0013	0.0043	-0.0018
(period t+2)*IDEX Awarded	-0.13***	-0.54*	-0.12***	-0.013	-0.085	0.020	0.023	-0.042	0.000026	0.0044	-0.0023
(period t+3)*IDEX Awarded	-0.10***	-0.60*	-0.12***	-0.027***	-0.13*	0.046*	0.19***	0.15	-0.00084	0.0036	-0.00039
(period t+4)*IDEX Awarded	-0.032	-0.80**	-0.050*	-0.0094	-0.051	0.073***	0.20***	0.62***	0.0022	0.0099**	-0.0029
period t-5	-0.15***	-0.081	-0.031	-0.14***	-0.64***	0.085**	0.23***	0.36	0.00047	0.0028	0.013**
period t-4	-0.14***	-0.13	-0.14***	-0.11***	-0.55***	0.12***	0.29***	0.31	0.0026	0.017	0.010**
period t-3	0.0070	0.041	0.044	-0.049***	-0.22**	0.052**	0.14**	0.37*	0.0019	0.0096	0.0044
period t-2	-0.019	0.094	0.015	-0.046***	-0.16*	0.057**	0.029	0.42**	0.0050*	0.015*	0.0020
period t	0.12***	-0.14	0.054**	0.044***	0.26***	-0.085***	0.020	-0.19	0.0013	0.0056	-0.0041
period t+1	0.099**	-0.46*	0.0074	0.055***	0.28***	-0.055**	-0.048	0.038	0.0078**	-0.0059	-0.010***
period t+2	0.21***	-0.24	0.077**	0.096***	0.22*	-0.11***	-0.18**	-0.38*	0.0047	-0.000037	-0.011**

period t+3	0.18***	-0.44	0.023	0.14***	0.21	-0.11***	-0.24***	-0.75***	0.0085	-0.0042	-0.018***
period t+4	0.16***	-1.13***	-0.024	0.14***	0.063	-0.15***	-0.40***	-0.55*	0.0062	-0.051**	-0.017**
Res. Seniority	0.11***	0.48***	0.080***	0.0035	0.096***	0.050***	0.15***	0.26***	0.00090	0.020***	0.0064***
Res. Seniority ^2	-0.004***	-0.012***	-0.0023***	-0.00073***	-0.0030***	-0.00043***	0.00020	-0.00075	-0.000044***	-0.00063***	-0.00010***
Constant	1.84***	4.75***	0.70***	0.58***	3.29***	-0.15	-0.87***	0.99	0.018	0.0016	-0.050**
Observations	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470	329,470
R-squared	0.009	0.002	0.010	0.008	0.012	0.005	0.016	0.007	0.000	0.008	0.001
Number of researchers	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947	32,947
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the whole sample of 32,947 active French researchers. Leads, lags, and period dummies are included in the regressions. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variables *IDEX Applicant* and *IDEX Awarded*. However, these dummies are collinear to the researcher fixed effects and are not included in the regressions.

Table L2. Joint significance test of the leads and lags.

Within Lab Co-authors	
Leads	Lags
(1) (period t-5)*IDEX Applicant = 0	(1) (period t)*IDEX Applicant
(2) (period t-4)*IDEX Applicant = 0	(2) (period t+1)*IDEX Applicant
(3) (period t-3)*IDEX Applicant = 0	(3) (period t+2)*IDEX Applicant
(4) (period t-2)*IDEX Applicant = 0	(4) (period t+3)*IDEX Applicant
(5) (period t-1)*IDEX Applicant = 0	(5) (period t+4)*IDEX Applicant
F (4, 32946) = 1.37	F (5, 32946) = 3.63
Prob > F = 0.2431	Prob > F = 0.0028

Table L3. Direct effect of being awarded IDEX on French researchers' outcomes, using the CEM sample. OLS estimates including leads and lags.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pubs	Citation Weighted Pubs	Interdiscipl. Pubs	Interdiscipl. Co-authors	Within Lab Co-authors	Within University Co-authors	National Co-authors	International Co-authors	At least one Patent	At least one Ph.D. Student	At least one ANR Grant
(period t-5)*IDEX Awarded	-0.024	-0.095	-0.016	-0.00050	-0.024	0.018	-0.013	0.0020	0.0023	-0.0012	-0.0017
(period t-4)*IDEX Awarded	-0.0084	-0.037	-0.042*	0.0022	-0.064	0.011	-0.051	-0.068	0.00017	-0.0087	0.0015
(period t-3)*IDEX Awarded	-0.051	-0.080	-0.045*	-0.019	-0.20***	-0.029	-0.016	-0.012	0.0045*	-0.0018	-0.0010
(period t-2)*IDEX Awarded	-0.026	-0.093	-0.027	-0.0070	-0.087	0.037*	0.0080	-0.12*	0.00034	-0.00034	0.00017
(period t-1)*IDEX Awarded	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
(period t)*IDEX Awarded	-0.013	0.046	0.0072	-0.033**	-0.033	0.067***	0.11***	0.24***	-0.00067	-0.0057	0.0018
(period t+1)*IDEX Awarded	-0.0097	0.099	0.0084	-0.0075	0.023	0.044*	0.079*	0.28***	0.00017	-0.0045	0.00050
(period t+2)*IDEX Awarded	-0.099***	-0.047	-0.044*	-0.033**	-0.21***	0.047*	-0.026	0.12	0.00084	0.0042	-0.0017
(period t+3)*IDEX Awarded	-0.043	0.11	-0.025	-0.022	-0.15*	0.082***	0.11**	0.16	-0.00050	-0.00084	0.0027
(period t+4)*IDEX Awarded	-0.026	0.061	0.0028	-0.019	-0.091	0.091***	0.11*	0.14	0.0035	-0.0054	0.0025
period t-5	-0.033	-0.0011	-0.038**	-0.038***	-0.32***	-0.057***	-0.17***	-0.093**	-0.0044**	-0.0084	-0.00084
period t-4	-0.10***	-0.24***	-0.11***	-0.037***	-0.33***	-0.065***	-0.15***	-0.078*	-0.0017	0.0027	-0.0013
period t-3	0.079***	0.035	0.054***	-0.0067	-0.024	0.00067	-0.11***	0.020	-0.0034*	0.0052	-0.0012
period t-2	0.052**	0.11	0.034**	0.00084	-0.029	-0.040***	-0.061***	0.042	0.00034	0.0077	-0.00084
period t	0.11***	0.57***	0.071***	0.061***	0.30***	0.017	0.050*	0.23***	0.000	-0.0032	0.0069***
period t+1	0.11***	0.50***	0.073***	0.050***	0.33***	0.052***	0.16***	0.32***	0.0027	0.0057	0.0054***
period t+2	0.19***	0.63***	0.12***	0.076***	0.43***	0.060***	0.29***	0.51***	0.000	-0.0069	0.0034**
period t+3	0.14***	0.59***	0.095***	0.076***	0.40***	0.055***	0.24***	0.60***	0.0027	-0.0055	0.0020
period t+4	0.073***	0.093	0.051***	0.075***	0.28***	0.058***	0.26***	0.68***	-0.0012	-0.038***	0.0012
Constant	1.42***	3.01***	0.81***	0.33***	2.31***	0.25***	0.60***	1.16***	0.014***	0.12***	0.0058***
Observations	119,380	119,380	119,380	119,380	119,380	119,380	119,380	119,380	119,380	119,380	119,380
R-squared	0.004	0.003	0.005	0.006	0.009	0.004	0.012	0.008	0.000	0.002	0.002
Number of researchers	11,938	11,938	11,938	11,938	11,938	11,938	11,938	11,938	11,938	11,938	11,938
Researcher F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This analysis is conducted on the sample of researchers obtained using the CEM matching. It includes 119,380 researcher-year pairs, 11,938 researchers observed for ten years each, divided into 5,969 awarded researchers and 5,969 applicant but not awarded similar researchers. Leads, lags, and period dummies are included in the regressions. Standard errors are clustered around the researcher. Significance levels at ***p<0.01, **p<0.05, *p<0.1. The correct estimation of the difference-in-differences model should have also included the non-interacted variable *IDEX Awarded*. However, this dummy is collinear to the researcher fixed effects and is not included in the regressions.

Table L4. Joint significance test of the leads and lags.

National Co-authors	
Leads	Lags
(1) (period t-5)*IDEX Applicant = 0	(1) (period t)*IDEX Applicant
(2) (period t-4)*IDEX Applicant = 0	(2) (period t+1)*IDEX Applicant
(3) (period t-3)*IDEX Applicant = 0	(3) (period t+2)*IDEX Applicant
(4) (period t-2)*IDEX Applicant = 0	(4) (period t+3)*IDEX Applicant
(5) (period t-1)*IDEX Applicant = 0	(5) (period t+4)*IDEX Applicant
F (4, 11937) = 0.9	F (5, 11937) = 2.86
Prob > F = 0.4646	Prob > F = 0.0140
International Co-authors	
Leads	Lags
(1) (period t-5)*IDEX Applicant = 0	(1) (period t)*IDEX Applicant
(2) (period t-4)*IDEX Applicant = 0	(2) (period t+1)*IDEX Applicant
(3) (period t-3)*IDEX Applicant = 0	(3) (period t+2)*IDEX Applicant
(4) (period t-2)*IDEX Applicant = 0	(4) (period t+3)*IDEX Applicant
(5) (period t-1)*IDEX Applicant = 0	(5) (period t+4)*IDEX Applicant
F (4, 11937) = 1.53	F (5, 11937) = 2.66
Prob > F = 0.1893	Prob > F = 0.0208

APPENDIX M

This appendix describes how we assign a field of research to each researcher's co-author when calculating the variable *Interdisciplinary Co-authors*.

We refer to *SCOPUS's All Science Journal Classification scheme (ASJC)* which is used by SCOPUS to classify journals under subject areas. We assign each researcher's co-author to the most frequent field of research of the journals where she published her research work during the period 2006-2015.

In doing so, we reaggregate the SCOPUS subject areas into five fields of research as described in Table M1. Finally, we assign one of these five fields to each researcher's co-author.

Table M1. Reaggregation scheme of SCOPUS subject areas into fields of research.

Field of research	***SCOPUS ASJC Subject Area Classifications***
Life sciences	Agricultural and Biological Sciences
	Biochemistry, Genetics and Molecular Biology
	Immunology and Microbiology
	Neuroscience
	Pharmacology, Toxicology and Pharmaceutics
Medicine	Medicine
	Nursing
	Veterinary
	Dentistry
	Health Professions
Mathematics	Mathematics
	Computer Science
Engineering	Engineering
	Chemical Engineering
	Energy
Physics	Physics and Astronomy
	Earth and Planetary Sciences
	Environmental Science
	Material Science
	Chemistry

APPENDIX N

Table N1 lists the 26 French universities that were the potential applicants of IDEX, with their status concerning the first round of IDEX funding.

Table N1. List of French universities and their status concerning the 1st round of IDEX funding.

HE&R Clusters	*IDEX Applicant*	*IDEX Awarded*
Université de Bordeaux	Yes	Yes
Université Grenoble Alpes	Yes	
Normandie Université		
PSL Université Paris Sciences et Lettres	Yes	Yes
Université Paris Saclay	Yes	Yes
Université Paris Est	Yes	
Université Paris Lumières		
Université Paris Seine		
Sorbonne-Paris-Cité	Yes	Yes
Sorbonne Université	Yes	Yes
HESAM - Hautes Ecoles Sorbonne Arts et Métiers Université	Yes	
Université de Bourgogne Franche-Comté	Yes	
Université de Lyon	Yes	
Université de Toulouse	Yes	Yes
Université Côte d'Azur		
Université Bretagne-Loire	Yes	
Université confédérale Léonard de Vinci		
Université Centre Val de Loire		
Université de Strasbourg	Yes	Yes
Université d'Aix-Marseille	Yes	Yes
Université Clermont Auvergne		
Université de Lorraine	Yes	
Université de Picardie Jules Verne		
Université de Reims		
Languedoc-Roussillon Universités	Yes	
Université Lille Nord de France	Yes	

CHAPTER 3

Is grant-funded research more impactful than block-funded research? Evidence from France

3.1 Introduction

Europe is experiencing tension between two complementary models for science funding: the competitive grant funding model and the more traditional block funding model (Geuna, 2001; Shibayama, 2011; Stephan, 2012; Lewis, 2015; Veugelers et al., 2022; The Economist, 2021). The competitive grant funding model echoes the one adopted historically in the US and consists of public agencies assigning funds to researchers through a peer-review process evaluating project proposals. The block funding model echoes the one historically adopted by European countries and consists of a steady stream of funding, allocated either incrementally or on a formula basis, addressed directly to universities and research institutions.

Although scientific literature has already shown that funding is one of the levers for knowledge creation (Martin, 2003; Defazio et al., 2009; Heyard and Hottenrott, 2021), the literature analyzing the effectiveness of different funding models is still scant (Geuna and Martin, 2003; Stephan, 2012; Wang et al., 2018). Nonetheless, studying funding models is crucial because it concerns the effectiveness of governments' spending decisions for advancing science and technology, and, ultimately, fostering countries' economic growth (Lane and Bertuzzi, 2011; Oancea, 2019; OECD, 2019).

The question of whether the impact of knowledge is different when produced through the competitive funding system or under the block funding approach is still unanswered. This paper aims to contribute to the literature by comparing the impact of the research produced through competitive grant funding and institutional block funding.

The two funding models differ substantially. The main advantage of block funding is to provide researchers with more stability and autonomy, encouraging them to undertake more risky research avenues. The main drawback of block funding is that it is often provided independently from the researchers' past scientific outcomes, decreasing the incentives for researchers to remain productive over time (Stephan, 2012). Unlike block funding, the grant funding model encourages competitiveness among scientists providing them an incentive to remain productive throughout their careers. Moreover, being funds distributed based on proposed projects, grant funding allows young non-established scholars to gather resources to compete with seniors. The main criticism addressed to the grant funding model is that it encourages producing short-term, low-risk, and measurable results due to the risk-averse behavior of the funding agencies. Another drawback of competitive grants is that they divert scientists' attention by subtracting hours from their core research activities to craft research proposals (Alberts, 2010; Stephan, 2012; Stephan et al., 2017; Veugelers et al., 2022; Franzoni et al., 2022). According to these characteristics, block funding and grant funding are expected to have different effects on the impact of the resulting research in terms of extent and time horizon.

To compare the effect of competitive grant funding and block funding, we identify the articles supported by grants distributed by the *Agence Nationale de la recherche* (ANR), the French main funding agency, and block-funded articles authored by researchers affiliated with French institutions. We rely on scientific articles' acknowledgments to identify grant- and block-funded articles published between 2009 and 2013. Hence, we implement probabilistic matching to compare 6,441 grant-funded articles with 6,441 similar block-funded articles. To assess the articles' impact, we count the yearly citations they received in the short and long run. In other words, if t is the article's publication date, we consider the citations in the short run as those received between t and $t+2$, and the citations in the long run as those received between $t+3$ and $t+5$.

Different from previous studies that used selected publication samples or disciplines (Tonta and Akbulut, 2020; Álvarez-Bornstein and Bordons, 2021), we identify the entire set of publications resulting from grants awarded by a national funding agency and those resulting from institutional block funding in a big European country. We isolate articles supported exclusively by ANR grants as a competitive funding source and articles supported exclusively by block funding. Then, we compare the impact of the two science funding models using a Propensity Score Matching approach as identification strategy (Rubin, 2001; Dehejia and Wahba, 2002). This technique allows us to mitigate the selection bias that arises when comparing grant-funded research with (systematically different) block-funded research by relying on a comparison of similar publications according to several observable authors' and articles' characteristics.

Our main finding is that articles resulting from competitive grants receive more citations than articles resulting from block funding in the long run. Specifically, grant-funded articles receive, on average, about 7% more citations than block-funded articles in the long run. When breaking down our analysis into four different research fields, we find that articles in Life sciences and Medicine, Engineering, and Physical sciences are more impactful in the long run when supported by competitive grants. They receive about 15%, 14%, and 8% more citations than the articles supported by block funding, respectively. Articles in Engineering are also more cited in the short run when resulting from competitive grants (+8%). Articles in Mathematics follow a different pattern: when supported by grant funding, they are less impactful in the short run (-13%), while there is no statistical difference between competitive grants and block funding in the long run.

3.2 Studying the relationship between funding and research impact

The empirical literature aiming to quantify the impact of research funding is still limited and far from reaching a consensus (Arora et al., 2000; Arora and Gambardella, 2005; Jacob and Lefgren, 2011). The vast majority of existing studies focus on competitive grants awarded to individual researchers, neglecting the impact of block funding. Keeping track of block funding investments is challenging since the spending decisions are often left to the discretion of research institutions.

When looking at competitive grants, prior literature focusing on researchers suggests that funding might have a positive effect by increasing researchers' productivity significantly (Jacob and Lefgren, 2011; Hottenrott and Lawson, 2017; Carayol and Lanoe, 2018; Gush et al., 2018; Heyard and Hottenrott, 2021). Studies at the publication level across fields report a positive impact of competitive grants on funded publications' citations received and the prestige of the journals where they are published (Campbell et al., 2010; Zhao, 2010; Wang and Shapira, 2015; Yan et al. 2018; Álvarez-Bornstein and Bordons, 2021). However, part of the existing literature shows the ambiguity of the relationship between funding and research outcome. Arora and Gambardella (2005) find a modest positive impact of the NSF grants only on young researchers' productivity in economics. Mariethoz et al. (2021) claim no correlation between researchers' publication and citation records and grant funding in Geosciences. Also, the effect of competitive grants on resulting publications' impact seems to be very modest for some specific journals (Rigby, 2013), in some specific fields (Haslam et al., 2008), or when focusing on particular funding agencies (Langfeldt et al., 2015) or countries (Tonta and Akbulut, 2020).

We expect to find in our analysis a positive effect of competitive grants on publications' number of citations that may vary over time due to two possible mechanisms. On the one hand, funding agencies claim to support breakthrough research (ANR, 2020; Franzoni et al., 2022). Along this line, the selection process conducted by the ANR funding agency is expected to recognize disruptive research projects that will lead to research articles more likely to be cited in the long run (Lewison and Dawson, 1998; Wang et al., 2017). The likelihood of producing disruptive research increases due to the additional resources provided by the competitive grants that bear the cost of state-of-the-art equipment, access to data, and an additional workforce in the lab (Katz and Martin, 1997). Part of these resources is contractually assigned to favor better dissemination of research results to foster the articles' scientific impact in the long run[74]. On the other hand, a stream of recent literature is skeptical about funding agencies' statements and highlights how funding agencies tend instead to be risk-averse in their decisions by selecting safe projects that ensure concrete results in the short-term (Stephan et al., 2017; Veugelers et al., 2022; Franzoni et al., 2022). Therefore, the impact of articles resulting from competitive grants might be greater than that of block-funded articles either in the short or long run.

We also expect that block funding brings advantages to researchers. One reason is that block funding provides researchers with more financial stability, autonomy, and flexibility. Researchers who rely on block funding should be encouraged to pursue risky research with long-term benefits. Moreover, block funding does not bound research ideas within the framework of a submitted research proposal like grant funding. Researchers funded with block funding can quickly adapt their research line to more promising ideas. On the contrary, grant-funded researchers are obliged to pursue their initial idea and must deliver results related to the proposed project. If the project does not develop according to the initial researchers' expectations, the results might be less impactful. For these reasons, block-funded publications might have a higher impact than grant-funded publications either in the long run, if scientists engage in risky research, or in the short run, if researchers tend to (re)shape their research according to the most promising results to obtain immediate recognition from the scientific community (Wei et al., 2013).

Finally, we expect research funding to have different effects on publications' impact across research fields. Research fields differ in research methods, teamwork approach, international vocation, and how they use the funding. Researchers in equipment–based fields, such as Engineering or Physical sciences, may benefit from grants' additional resources by purchasing the state-of-the-art equipment needed to carry out impactful research. Moreover, collaboration-oriented research fields may need additional funding to cultivate broad collaborations, hence competitive grants supporting mobility may favor research outcomes. Therefore, we expect to find

[74] See a 2022 ANR *Generic Call for Proposals*, page 20. A part of the budget is foreseen for a "Strategy for disseminating and exploiting results; promoting scientific, technical and industrial knowledge". Website: https://anr.fr/fileadmin/aap/2022/aapg-2022-v1.1a-en.pdf.

a greater impact of competitive grants than block funding in the long run in equipment-based and collaboration-oriented research fields. On the contrary, fundamental research and less collaboration-oriented research, such as Mathematics, might benefit less from competitive grants. Thus, in the case of Mathematics, we expect block funding and competitive grants not to have a different effect on the impact of the resulting articles.

Previous studies show gaps in large-scale comparison of the effects of competitive grants and block funding on the impact of the research produced. We identified three main reasons for explaining this. First, it is generally difficult to identify block-funded publications. Second, studies at the researcher level struggle to isolate the effect of a grant if other grants are available to a researcher simultaneously. Last, studies at the article level compare papers acknowledging different funding sources without considering the potential selection problem due to the fact that different types of grants might support articles with different characteristics correlated to the articles' impact.

We conduct our empirical analysis in France. France is one of the European countries that most rely on public investments to support research and it differentiates from the other European countries for not having reduced the level of block funding over the last fifteen years. Around 90% of researchers with a French affiliation benefit from government block funding through a monthly salary paid by their university of affiliation or by a national public research organization (PRO)[75] (OECD, 2019; Pommier et al., 2022). Concerning competitive grants, France launched its national funding agency in 2005, namely *Agence Nationale de la recherche* (ANR). ANR is currently the most important French funding agency. Since 2006, it has distributed around 1,100 individual grants per year. In 2019, it awarded 1,157 research projects with an average 400 thousand euros budget per project. Applications from all disciplines are eligible. ANR is also the French government's operator for the IDEX competitive funding program addressed to French universities. The main priority of the ANR agency is to "promote research in all its forms […] on the principle of peer review based on scientific excellence"[76].

[75] The largest public research organization in France is the *Centre national de la recherche scientifique* (CNRS).
[76] Website: https://anr.fr/fileadmin/documents/2021/ANR-RA2020-en.pdf, page 04.

3.3 Data and methodology

3.3.1 Data

We rely on the funding acknowledgment information reported in scientific articles to identify block-funded and grant-funded publications (Rigby, 2011; Gok et al. 2016; Grassano et al., 2016). We retrieve the acknowledgment information from the *Web of Science* (WOS, Clarivate) bibliometric dataset[77] while we use *Microsoft Academic Graph* (MAG) citation data to assess the publication impact.

We construct our universe of publications by collecting all the 481,536 publications issued between 2009 and 2013, having at least one author affiliated with a French institution and a *Digital Object Identifier* (DOI). We choose 2009 as the starting date of our analysis because WOS acknowledgment data are reliable only starting from that date (Mongeon and Paul-Hus, 2016; Mejia and Kajikawa, 2018), while we choose 2013 as the end date because MAG citation data are not reliable after 2018 and, for the articles published in 2013, we need a 6-year forward citation window to evaluate their impact. We focus on scientific articles and exclude other types of publications, such as reviews or book chapters. Moreover, we limit our analysis to publications written in English because WOS collects acknowledgments only for those publications. Finally, we exclude journals in Social Sciences and Humanities because, for these disciplines, WOS reports acknowledgments only starting from 2015 (Álvarez-Bornstein et al., 2017; Liu et al., 2020). After applying all the previously mentioned constraints, we ended up with a sample of 283,873 publications.

We use the information from Elsevier's SCOPUS database to retrieve articles' and authors' characteristics. We add information on the authors' gender, not available in SCOPUS, by matching the authors' given names with French[78] and international gender-name datasets[79]. To characterize the reputation of the authors' affiliations, we use the *QS university ranking*[80]. Our sample of 283,873 publications cannot be promptly matched with the complementary information retrieved from these databases. For instance, when we attribute gender to authors by matching the authors'

[77] Web of Science (WOS) provides two acknowledgment fields, one reporting the raw text as written in the paper and the other is an artificial field that already extracted the names of the funding organizations from the raw text through an algorithm. The WOS algorithm does not seem to be accurate, thus we rely on the raw text of the acknowledgments.
[78] Website: https://www.data.gouv.fr/fr/datasets/liste-de-prenoms/
[79] Authors' names non-matched with the French dataset are matched with the *U.S. Census Bureau* gender-name dataset and the WIPO gender-name dataset (website: https://www.wipo.int/publications/en/details.jsp?id=4125).
[80] Website: https://www.topuniversities.com

names with the gender-name datasets, we cannot attribute gender to all the authors for 84,943 articles. In this case, we remove the corresponding articles from our dataset. After matching articles' and authors' characteristics, we end up with a sample of 195,435 articles.

3.3.2 Identifying grant-funded and block-funded articles

From our sample of 195,435 articles, we selected all the 23,950 publications acknowledging at least one ANR grant among the funding sources (12.25% of the publications in our sample). To avoid mixing the effects of several funding sources contributing to the research outcome described in an article, we limit our sample of grant-funded articles to the 6,441 publications that report ANR grants as the sole competitive funding source. In other words, we excluded the publications acknowledging an ANR grant along with other non-ANR competitive grant funding[81].

To identify publications treated with block funding, we considered articles that do not report any acknowledgment. The logic is that if no grant acknowledgment is reported in an article, the research outcome described in the article is likely to result from block funding[82]. Among the 195,435 publications in our sample, we identified 76,615 articles with no acknowledgments. This figure corresponds to 39.2% of our sample, in line with previous studies (Grassano et al., 2016).

3.3.3 Propensity Score Matching

Our analysis aims at comparing the impact of articles supported by ANR competitive grants with the impact of articles resulting from block funding. A simple comparison between the number of citations received by grant-funded and block-funded articles is likely to be affected by a selection problem (Jaffe, 2002). Indeed, grant-funded and block-funded articles might systematically differ in other aspects than the funding source. The articles' or authors' characteristics might relate both to the likelihood of observing grant-funded or block-funded research and to the articles' impact. These characteristics reflect the collaborative behaviors, team composition, and stock of knowledge that influence both funding attractiveness and publication impact (Wuchty et al., 2007; Ebadi and Schiffauerova, 2015a; Mukherjee et al. 2017; Bol et al., 2018; Bianchini et al., 2022). Therefore, we adopt a Propensity Score Matching (PSM) procedure

[81] See the robustness check of section 3.6.5 for an empirical analysis in which we consider as ANR-funded articles also the articles supported by additional competitive grants other than ANR.
[82] The ANR funding agency provides specific guidelines to French universities and their researchers on how to report acknowledgments. We queried a subset of French universities' researchers who confirmed they were aware of the acknowledgment guidelines.

that relies on the *nearest neighbor* approach to mitigate the potential selection bias. This approach compares each publication funded by an ANR grant with a similar block-funded publication. The similarity between publications is assessed by a probabilistic score based on the articles' and authors' observable characteristics.

To retrieve for each ANR grant-funded article a similar block-funded article, we estimate a propensity score equation calculating the predicted probability that an article is supported by an ANR grant based on the article's and its authors' observable characteristics. Specifically, we run a logit regression where the left-hand-side variable is represented by a dummy variable that equals one if the publication acknowledges an ANR grant as the sole supporting competitive funding and zero if it does not report any acknowledgment (i.e., is block-funded). The right-hand side variable is represented by a vector containing the variables measuring the article's and authors' characteristics. We include three dummy variables characterizing the co-authorship behavior. The dummy *Single-author article* equals one when the article has only one author. The dummy *Multi-author article (from 2 to 4 authors)* equals one when the article has between two and four co-authors, while the dummy *Multi-author article (more than 4 authors)* equals one when the article has more than four co-authors. To account for international collaborations, we include the dummy variable *At least one international author* that equals one if there is at least one international author among the article's authors. We define international authors as those reporting only non-French affiliations in the focal article. To account for the team gender composition, we add the dummy variable *At least one female author* that equals one if there is at least one female among the article's authors. Moreover, we identify the article's authors affiliated with top-ranked universities with the dummy variable *At least one top-affiliate author* that equal to one if there is at least one author affiliated with a top-ranked university among the article's authors, zero otherwise. To identify the top-ranked universities, we rely on the *QS ranking*[83]. Specifically, we identify the top-ten-ranked universities in France and the top-fifty-ranked universities worldwide. Then, we include the dummy variable *Multiple affiliations* to account for the geographical dispersion of the authors. *Multiple affiliations* equals one if the total number of distinct affiliations reported in the focal article is greater than one. We also include four dummy variables representing the quartiles of the articles' backward citation distribution. Specifically, we created the dummy *Backward citations*

[83] https://www.topuniversities.com/university-rankings. We gather the ranking information in 2020, however university ranking has minor variation over the years when considering top-universities.

Q1 that equals one if the number of article's backward citations belongs to the first quartile of the distribution of our sample of articles in the same publication year. With the same logic, we created the dummy variables *Backward citations Q2*, *Backward citations Q3*, and *Backward citations Q4*, for the second, third, and fourth quartiles, respectively. To account for the heterogeneity in the publication behaviors across fields of study, we add a set of dummy variables that classify the articles in four fields of study, according to the journals where they have been published[84]. The dummy *Mathematics* equals one if the article is published in a Mathematical journal, the dummy *Engineering* equals one if the article is published in an Engineering journal, the dummy *Physical sciences* equals one if the article is published in a journal classified in Physical sciences, and the dummy *Life sciences and Medicine* equals one if the article is published in a journal classified in Life sciences or Medicine. Finally, to control for the cohort publication effect, we include the variable *Year of publication* which specifies the year when the article is published.

All the variables are listed in Table 12, along with a brief description of how we calculated them.

[84] See Appendix V for a detailed explanation of the research field classification.

Table 12. List of variables used to predict the probability of observing an ANR grant-funded article.

Articles' and authors' characteristics	Variable description
Single-author article	Dummy variable that equals one if the article has only one author.
Multi-author article (from 2 to 4 authors)	Dummy variable that equals one if the article has two to four authors.
Multi-author article (more than 4 authors)	Dummy variable that equals one if the article has more than four authors.
At least one international author	Dummy variable that equals one if the article has at least one international author.
At least one female author	Dummy variable that equals one if the article has at least one female author.
At least one top-affiliate author	Dummy variable that equals one if at least one article's author is affiliated with a university ranked in the top ten in France or the top 50 worldwide according to the QS ranking.
Multiple affiliations	Dummy variable that equals one if the affiliations of the article's authors are more than one at the publication date.
Backward citations Q1	Dummy variable that equals one if the number of article's backward citations belongs to the first quartile of the backward citation distribution of our sample of articles in the same publication year.
Backward citations Q2	Dummy variable that equals one if the number of article's backward citations belongs to the second quartile of the backward citation distribution of our sample of articles in the same publication year.
Backward citations Q3	Dummy variable that equals one if the number of article's backward citations belongs to the third quartile of the backward citation distribution of our sample of articles in the same publication year.
Backward citations Q4	Dummy variable that equals one if the number of article's backward citations belongs to the fourth quartile of the backward citation distribution of our sample of articles in the same publication year.
Life sciences and Medicine	Dummy variable that equals one if the article is published in a journal classified in Life sciences or Medicine.
Mathematics	Dummy variable that equals one if the article is published in a journal classified in Mathematics.
Engineering	Dummy variable that equals one if the article is published in a journal classified in Engineering.
Physical sciences	Dummy variable that equals one if the article is published in a journal classified in Physical sciences.
Year of publication	The publication year of the article.

Table 13 shows the average marginal effects calculated from estimated logit coefficients. We find that all the articles' and authors' characteristics are associated with the likelihood that an ANR competitive grant supports a publication. Specifically, we find that the increasing number of the article's authors is associated with a higher probability that the article is supported by an ANR competitive grant. Articles with more than four authors are 5.8% more likely to be supported by competitive grants than single-author articles. Interestingly, we find that the presence of an international author among the article's authors is associated with a 7.3% lower probability that the article is supported by ANR competitive grants. The presence of a female author is associated with a slightly lower probability of being supported by ANR grants (-0.71%), while the presence of an author affiliated with a top-ranked university is associated with a 2% higher probability that the article is supported by a competitive grant. More than one affiliation listed in the article is associated with a 1.1% higher probability of obtaining ANR support. Concerning the article's characteristics, increasing the number of backward citations is associated with a higher probability that an ANR grant supports the article. Articles belonging to the fourth quartile of the backward citations distribution are 6% more likely to be supported by ANR grants than articles belonging to the first quartile. Finally, articles published in Mathematical and Physical journals are more likely to be supported by ANR competitive grants than articles in Life sciences, Medicine, and Engineering, as well as articles published in recent years.

Using the estimates reported in Table 13, we predict the probability of an article being supported by an ANR grant. According to these predictions, for each article funded by ANR, we match an article drawn from the sample of 76,615 block-funded articles with the highest propensity score similarity following the *nearest neighbor* Propensity Score Matching procedure. Figure 5 reports the histograms of the density of propensity scores for publications before and after matching. We observe that after the matching, the distribution of the propensity scores is very similar between grant-funded and block-funded articles.

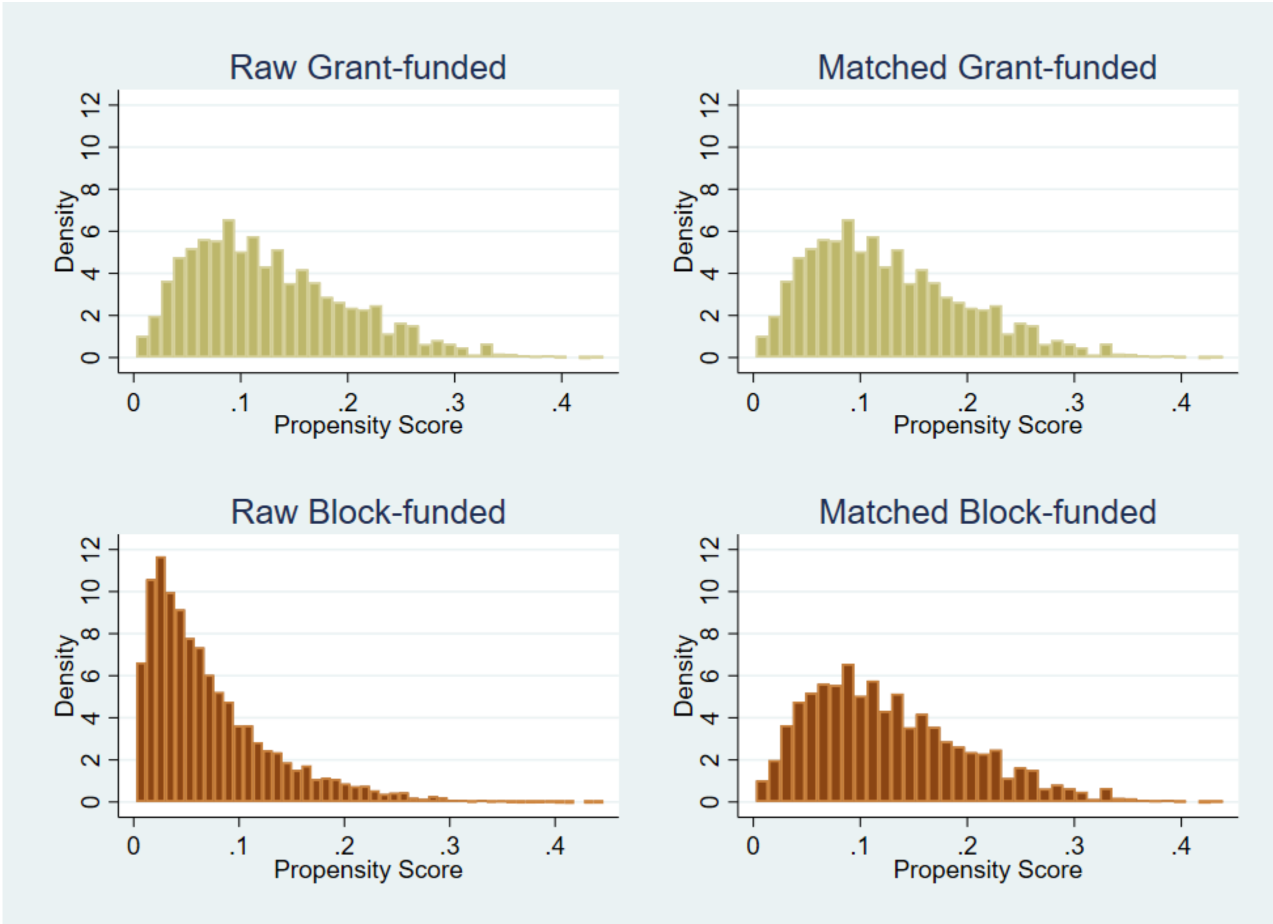
Finally, Table 14 compares the characteristics of the grant-funded articles with those of the block-funded articles for the raw sample before the PSM matching (columns 1 and 2) and the sample of similar publications obtained after applying the PSM matching procedure (columns 3 and 4). Columns 3 and 4 show that we matched the sample of grant-funded articles with a sample of block-funded articles with statistically equivalent articles' and authors' characteristics.

Table 13. Average marginal effects of the article's probability of being supported by an ANR grant.

	(1) **Grant-funded (=1) versus Block-funded (=0)**
Single-author article	Ref.
Multi-author article (from 2 to 4 authors)	0.044*** (0.0037)
Multi-author article (more than 4 authors)	0.058*** (0.0042)
At least one international author	-0.073*** (0.0026)
At least one female author	-0.0071*** (0.0021)
At least one top-affiliate author	0.020*** (0.0019)
Multiple affiliations	0.011*** (0.0022)
Backward citations Q1	Ref.
Backward citations Q2	0.026*** (0.0030)
Backward citations Q3	0.042*** (0.0029)
Backward citations Q4	0.060*** (0.0028)
Life sciences and Medicine	Ref.
Mathematics	0.070*** (0.0024)
Engineering	-0.0052** (0.0022)
Physical sciences	0.068*** (0.0020)
Year of publication	0.014*** (0.00064)
Pseudo R2	0.0858
Number of articles	83,056

NOTE: Average marginal effects are calculated from estimated logit coefficients. Standard errors in parentheses. Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Figure 5. Histograms of the density of propensity scores before (Raw) and after matching (Matched).



NOTE: Distribution of the propensity scores for grant-funded publications (top part, in light brown) and block-funded publications (bottom part, in dark brown), before (left part) and after (right part) the *nearest neighbor* Propensity Score Matching.

Table 14. Means of the articles' and authors' observable characteristics for grant- and block-funded publications, before (Columns 1 and 2) and after matching (Columns 3 and 4).

	Sample before PSM			PSM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	6,441	76,615		6,441	6,441	
Covariates						
Single-author article	0.075***	0.109	0.000	0.075	0.075	0.973
Multi-author article (from 2 to 4 authors)	0.539***	0.483	0.000	0.539	0.549	0.265
Multi-author article (more than 4 authors)	0.386***	0.408	0.001	0.386	0.376	0.246
At least one international author	0.153***	0.279	0.000	0.153	0.155	0.714
At least one female author	0.596***	0.628	0.000	0.596	0.599	0.788
At least one top-affiliate author	0.399***	0.315	0.000	0.399	0.400	0.928
Multiple affiliations	0.693*	0.682	0.058	0.693	0.692	0.924
Backward citations Q1	0.145***	0.269	0.000	0.145	0.143	0.861
Backward citations Q2	0.225***	0.249	0.000	0.225	0.225	0.950
Backward citations Q3	0.278***	0.241	0.000	0.278	0.280	0.829
Backward citations Q4	0.352***	0.240	0.000	0.352	0.352	0.985
Life sciences and Medicine	0.265***	0.492	0.000	0.265	0.257	0.316
Mathematics	0.280***	0.189	0.000	0.280	0.282	0.829
Engineering	0.231***	0.199	0.000	0.231	0.228	0.691
Physical sciences	0.577***	0.380	0.000	0.577	0.577	0.957
Year of publication	2011.31***	2010.90	0.000	2011.31	2011.30	0.661

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. PSM sample refers to the sample obtained after the *nearest neighbor* Propensity Score Matching.

3.4 Results

We estimate the impact of grant-funded articles by calculating the difference between the number of citations received by grant-funded articles and block-funded articles. Since we are interested in measuring the articles' impact in the short and long run, we count the number of yearly citations received by each article in the first 3 years after its publication to assess the short run impact (*t to t+2, citations*) and the number of yearly citations received by each article in the 3 years after the short run to assess the long run impact (*t+3 to t+5, citations*). We rely on t-tests to estimate the difference in the funding effects. T-tests are valid for any distribution in large samples, including highly non-normal distributions (Lumley et al. 2002; Tonta and Akbulut, 2020).

Table 15 shows that articles resulting from ANR competitive grants receive, on average, 0.580 more citations than articles resulting from block funding in the long run (*Grant-funded effect*). The value of 0.580 corresponds to a 6.93% higher number of citations of ANR-funded articles than block-funded articles (*Grant-funded relative effect*). This result is statistically significant at the 0.05 level. In the short run, we do not find any statistically significant difference between the impact of articles resulting from ANR competitive grants and those resulting from block funding.

Table 15. Funding effects on publications' number of citations received.

6,441 Grant-funded + 6,441 Block-funded	(1) **Short run** t to t+2, citations	(2) **Long run** t+3 to t+5, citations
Average citations Grant-funded articles (A)	6.403	8.943
Average citations Block-funded articles (B)	6.628	8.363
Grant-funded effect (A-B)	-0.225	0.580 **
Grant-funded relative effect (A-B)/B	-3.4%	+6.93% **
t-statistic	*-1.438*	*2.450*
p-value	*0.15*	*0.014*

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

3.5 Heterogeneity across research fields

To dig into a possible cross-field heterogeneity of the funding effect on articles' impact, we run separated Propensity Score Matching exercises for four research fields. Specifically, we analyze publications in Life sciences and Medicine, Mathematics, Engineering, and Physical sciences. We use the journal field where an article is published to classify it in a research field. The journal field classification that we consider is the one provided by SCOPUS[85]. To avoid ambiguities in the classification, we excluded articles published in multidisciplinary journals, such as *Nature* or *Science*, given the impossibility of precisely assigning a research field to these journals. Moreover, some journals (and consequently articles) can be assigned to more than one field. This leads us to obtain 17,444 articles (1,708*2+1,806*2+1,489*2+3,719*2) when summing the number of articles in each of the four fields. This number is higher than the one reported in Table 15 (12,882=6,441*2).

Table 16 shows that publications resulting from ANR competitive grants benefit from a higher impact in the long run than publications resulting from block funding in three fields of research over four, namely Life sciences and Medicine, Engineering, and Physical sciences. Life sciences and Medicine, and Engineering, show the largest citation gap between grant-funded and block-funded articles, +15.06% and +13.87%, respectively. In Physical sciences, the citation gap in favor of grant-funded publications equals +7.67%. Interestingly, in Mathematics, there is not a statistically significant difference between the effects of competitive grants and block funding on the articles' impact in the long run. Moreover, in the short run, articles in Mathematics resulting from ANR competitive grants receive, on average, 13.08% fewer citations than articles resulting from block funding. On the contrary, in the short run, grant-funded articles in Engineering receive 8.15% more citations than block-funded articles.

[85] See Appendix V for a detailed explanation of the research field classification.

Looking at the other research fields in the short run, Life sciences and Medicine, and Physical sciences, we do not find a significant citation gap between grant-funded articles and block-funded articles.

In Appendix O, we report the tables showing the logit estimates of the articles' probability of being supported by an ANR grant and the covariate balance tables pre- and post-matching for each of the four research fields we consider.

Table 16. Funding effects on publications' number of citations received by field of research.

	(1) **Short run** t to t+2, citations	(2) **Long run** t+3 to t+5, citations
Life sciences and Medicine (1,708 Grant-funded + 1,708 Block-funded)		
Average citations Grant-funded articles (A)	8.228	11.888
Average citations Block-funded articles (B)	8.338	10.332
Grant-funded effect (A-B)	-0.11	1.556 ***
Grant-funded relative effect (A-B)/B	-1.32%	+15.06% ***
t-statistic	*-0.333*	*2.931*
p-value	*0.74*	*0.003*
Mathematics (1,806 Grant-funded + 1,806 Block-funded)		
Average citations Grant-funded articles (A)	3.950	5.679
Average citations Block-funded articles (B)	4.545	5.946
Grant-funded effect (A-B)	-0.595 **	-0.267
Grant-funded relative effect (A-B)/B	-13.08% **	-4.5%
t-statistic	*-2.163*	*-0.639*
p-value	*0.031*	*0.52*
Engineering (1,489 Grant-funded + 1,489 Block-funded)		
Average citations Grant-funded articles (A)	7.126	10.462
Average citations Block-funded articles (B)	6.589	9.188
Grant-funded effect (A-B)	0.537 *	1.274 **
Grant-funded relative effect (A-B)/B	+8.15% *	+13.87% **
t-statistic	*1.959*	*2.521*
p-value	*0.0502*	*0.012*
Physical sciences (3,719 Grant-funded + 3,719 Block-funded)		
Average citations Grant-funded articles (A)	6.805	9.292
Average citations Block-funded articles (B)	6.770	8.630
Grant-funded effect (A-B)	0.035	0.662 **
Grant-funded relative effect (A-B)/B	+0.51%	+7.67% **
t-statistic	*0.169*	*2.227*
p-value	*0.87*	*0.026*

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. The SCOPUS classification of journals (and consequently articles) might assign one journal to multiple research fields. For this reason, we obtain an overall sample of 17,444 articles (1,708*2+1,806*2+1,489*2+3,719*2) when summing the number of articles in each of the four fields. This number is higher than the one reported in Table 15 (12,882=6,441*2).

3.6 Robustness checks

This section presents six robustness checks to compare the impact of articles resulting from competitive grants with the impact of articles resulting from block funding.

3.6.1 Calculating the yearly citation impact

In Appendix P, we disentangle the effects of block and grant funding on the articles' citations received in each of the six years after publication. Table P1 shows that grant-funded articles receive fewer citations than block-funded articles in the first year after publication (-30.85%). Grant-funded articles become more cited than block-funded articles from the third year after publication, remaining more cited in the following years.

3.6.2 Using SCOPUS citation data

In Appendix Q, we calculate the articles' impact using the SCOPUS average yearly citations instead of Microsoft Academic Graph annual citations. To calculate the average yearly citations, we retrieved the cumulated number of each article's citations in 2018 from SCOPUS and divided it by the number of years elapsed between the article's publication date and 2018. In doing so, we allow for a time window longer than 6 years for articles published before 2013, but we cannot distinguish between the short and long run. Table Q1 shows that grant-funded articles receive, on average, 4.40% more yearly citations than block-funded articles.

3.6.3 Including French authors' characteristics

ANR grants are the main competitive grants obtained by French researchers. One of the main priorities of the ANR funding agency is to support French research on "the principle of peer review based on scientific excellence[86]". Seeking scientific excellence might bias the evaluation process in favor of French applicant researchers who are highly reputed or experienced scientists. Moreover, existing literature has shown how scientists with high academic recognition are more likely to have easier access to new grants for research, and obtain further over-recognition of their publications by the scientific community (Merton, 1968; Allison and Stewart, 1974; Allison et al., 1982; Ebadi and Schiffauerova, 2015a). On the contrary, low-status scientists seem to experience fewer benefits from competitive grants that discourage their novel research (Wang et al., 2018). The existence of a relationship between funding

[86] *Le contrat d'objectifs et de performance entre l'État et l'Agence nationale de la recherche 2021-2025.* Source: https://anr.fr/fileadmin/documents/2021/ANR_COP_2021-2025.pdf

and researchers' seniority is well documented in the literature, but the findings are ambiguous. Competitive grants might have a positive impact on young researchers' productivity, but disfavor their novel ideas (Arora and Gambardella, 2005; Wang et al., 2018; Heyard and Hottenrott, 2021).

Although we have already included the prestige of the researchers' university of affiliation as a characteristic to match grant-funded and block-funded publications, we add in this robustness check three variables proxying for the French researchers' academic status. We include the dummy variable *At least a French star author* that equals one if there is at least one French star scientist among the article's authors. We define a French start scientist as a researcher affiliated with at least one French institution who belongs to the last quartile of the distribution of French researchers according to their cumulative stock of citation-weighted publications from 1990 to the publication year of the article she authored. Then, we add the dummy *At least a French senior author* that equals one if there is at least one French senior scientist among the article's authors. We define a French senior scientist as a researcher affiliated with at least one French institution and whose first publication took place more than 10 years before the publication year of the article belonging to our sample that she authored. Finally, we include the dummy variable *At least a French Ph.D. student author* that equals one if there is at least one French Ph.D. student among the article's authors, i.e., a French author who has published the article at the latest one year after her Ph.D. defense date. To identify the French Ph.D. students, we rely on the French repository of *Electronic Doctoral Theses* (EDT), from which we retrieved the list of all the students who have deposited a Ph.D. thesis in France between 2000 and 2020.

We decided to implement this exercise as a robustness check and not as the main analysis for two reasons. First, we are able to create these variables only for French authors due to data limitations. Second, we lose observations when including these three variables in the matching exercise due to an increasing imbalance of the covariates. We still believe these variables are relevant since the main goal of the ANR funding agency is to support French research, thus in the evaluation process, the French authors' characteristics are likely to be relevant. Moreover, when calculating the probability that a publication is supported by ANR grants (Table 13), we find that the presence of an international author reduces the likelihood of being supported by an ANR grant by 7.3%.

After the PSM procedure, we end up with 5,537 grant-funded articles matched with 5,537 similar block-funded articles. Appendix R reports the table of the average marginal effects of the articles' probability of being supported by an ANR grant (Table R1) and the covariate balance table pre- and post-matching (Table R2), obtained after adding the three new covariates. After the matching, 53.5% of both grant-funded and block-funded publications have *At least a French star author*, 91.8% have *At least a French senior author*, and 38% have *At least a French Ph.D. student author*.

Table 17 reports the difference in the funding effects on the number of articles' citations received in the short and long run. Results are consistent with those of the main analysis (Table 15). The magnitude of the grant-funded relative effect, in the long run, is slightly larger than that of the main analysis (+7.33% versus +6.93%).

Table 17. Funding effects on publications' number of citations received.

5,537 Grant-funded + 5,537 Block-funded	(1) **Short run** t to t+2, citations	(2) **Long run** t+3 to t+5, citations
Average citations Grant-funded articles (A)	6.338	8.758
Average citations Block-funded articles (B)	6. 392	8.160
Grant-funded effect (A-B)	-0.054	0.598 **
Grant-funded relative effect (A-B)/B	-0.84%	+7.33% **
t-statistic	*-0.318*	*2.196*
p-value	*0.75*	*0.028*

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

3.6.4 Exact matching of publication years and journals

In our main matching exercise (Table 14) we let articles published in different years and different journals be paired together. However, this possibility is limited by the fact that we included among our covariates the variables *Year of publication* and four variables identifying the discipline of the journals where the articles are published, that are *Life sciences and Medicine*, *Mathematics*, *Engineering*, and *Physical sciences.* The number of matched publications having different years is very low: it concerns 120 article pairs. Table 14 shows how the mean of the publication year of the matched grant-funded articles is almost the same as that of the block-funded articles (2011.31 vs 2011.30). Instead, for what concerning journals, there is much more variation since different journals might belong to the same discipline. In our matched sample, only 109 article pairs are published in the same journal, while 6,332 (98.3%) pairs are published in different journals. Imposing the same year and journal of publication are strong criteria allowing us to control for the cohort effects, the citation trends, and the journals' editorial policies in a given year.

In this robustness check, we propose two exercises where we run the *nearest neighbor* Propensity Score Matching conditionally on (i) pairing articles published in the same year and (ii) pairing articles published in the same year and in the same journal.

Table 18 reports the results of the PSM, including the exact matching of the *Year of publication*. The propensity scores are calculated as in our main analysis and the average marginal effects of the articles' probability of being supported by grants are the same as reported in Table 13. Imposing exact matching on the year of publication means that among the matches having the same year of publication, we choose those with the closest distance measured by the propensity scores. In Table S1 of Appendix S, we report the covariate balance table pre- and post-matching. Table S1 shows that, after the matching, the covariate *Year of publication* is perfectly balanced between grant-funded articles and block-funded articles (it equals 2011.314 for both groups). The results reported in Table 18 are largely consistent with those of our main analysis (Table 15). The magnitude of the coefficient obtained when implementing the exact match on the publication year is slightly larger than that of the main analysis, moving from a grant-funded relative effect in the long run of +6.93% (Table 15) to +7.05% (Table 18).

Table 19 reports the results of the PSM including the exact matching of the *Year of publication* and the *Journal of publication*. Also in this case, the propensity scores are the same as our main analysis, as well as the average marginal effects of the articles' probability of being supported by ANR grants. In Table S2 of Appendix S, we report the covariate balance table pre- and post-matching. Table S2 shows that, after matching, the average year of publication is 2011.133 for both grant-funded and block-funded articles. Imposing an exact matching on journals and years while keeping the covariates balanced forces us to lose several observations. We pass from 6,441 article pairs to 1,643 article pairs. This reduction is because many grant-funded articles have no block-funded articles published in the same journal and year. Losing so many articles makes it difficult to find statistically significant differences between the impact of grant-funded articles and that of block-funded articles. Table 19 shows no statistically significant differences both in the short run and long run. We can still see that the direction of the grant-funded relative effect is positive. The positive effect of grant-funded articles, in the long run, has a similar magnitude as that found in the main analysis of Table 15.

Table 18. Funding effects on publications' number of citations received.

6,441 Grant-funded + 6,441 Block-funded	(1) **Short run** t to t+2, citations	(2) **Long run** t+3 to t+5, citations
Average citations Grant-funded articles (A)	6.404	8.943
Average citations Block-funded articles (B)	6.624	8.354
Grant-funded effect (A-B)	-0.22	0.589 **
Grant-funded relative effect (A-B)/B	-3.32%	+7.05% **
t-statistic	*-1.403*	*2.494*
p-value	*0.16*	*0.013*

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Table 19. Funding effects on publications' number of citations received.

	(1)	(2)
	Short run	**Long run**
1,643 Grant-funded + 1,643 Block-funded	t to t+2, citations	t+3 to t+5, citations
Average citations Grant-funded articles (A)	6.839	9.447
Average citations Block-funded articles (B)	6.471	8.986
Grant-funded effect (A-B)	0.368	0.461
Grant-funded relative effect (A-B)/B	+5.68%	+5.13%
t-statistic	*1.241*	*1.005*
p-value	*0.21*	*0.31*

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

3.6.5 Including articles supported by competitive grants other than ANR grants

In our main analysis, we compare the impact of the publications resulting from competitive grants with that of publications resulting from block funding. To select publications supported by competitive grants, we look at those that solely acknowledge the ANR funding agency as a competitive funding source. In doing so, we exclude all the publications resulting from ANR grants along with other competitive grants. In this robustness check, we consider as ANR grant-funded all the articles benefitting from at least an ANR grant, allowing for the presence of other non-ANR competitive funding sources. Given our main result that ANR grant-funded publications receive, in the long run, more citations than block-funded publications, we expect that adding additional sources of competitive funding will further boost the grant-funded effect.

Our sample before matching consists of 23,950 grant-funded publications supported by at least an ANR grant and 76,615 block-funded publications. After running the *nearest neighbor* Propensity Score Matching, we end up with a sample composed of 21,381 grant-funded articles matched with 21,381 similar block-funded articles, given the set of covariates listed in Table 12. The grant-funded publications are supported, on average, by 3.46 competitive grants. The median value is 3 competitive grants. Table T1 in Appendix T reports the average marginal effects of the articles' probability of being supported by at least an ANR grant. Table T2 shows the covariate balance tables pre- and post-matching.

Table 20 reports the difference in the funding effects on the number of citations received in the short and long run. We find that when considering articles supported by at least an ANR grant, grant-funded articles are significantly more impactful than block-funded articles both in the short and long run. Specifically, in the short run, the articles resulting from competitive grants receive 27.23% more citations than articles resulting from block-funding, while in the long run, they receive 25.56% citations more (Table 20, *Grant-funded relative effect*). These findings suggest that the presence of multiple competitive grants amplifies the impact of grant-funded publications.

Table 20. Funding effects on publications' number of citations received.

	(1) Short run	(2) Long run
21,381 Grant-funded + 21,381 Block-funded	t to t+2, citations	t+3 to t+5, citations
Average citations Grant-funded articles (A)	9.129	12.268
Average citations Block-funded articles (B)	7.175	9.771
Grant-funded effect (A-B)	1.954 ***	2.497 ***
Grant-funded relative effect (A-B)/B	+27.23% ***	+25.56% ***
t-statistic	*13.21*	*6.030*
p-value	*0.000*	*0.000*

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

3.6.6 Coarsened Exact Matching

The Propensity Score Matching approach is based on the estimation of the propensity score, that is the probability that an article of our sample is supported by an ANR competitive grant given the set of articles' and authors' characteristics listed in Table 12. In this robustness check, we implement the *Coarsened Exact Matching* (CEM) procedure (Azoulay, 2010; Iacus et al., 2012) to pair grant-funded publications with block-funded publications. Differently from the Propensity Score Matching, CEM is a nonparametric approach.

We keep the same set of covariates as in Table 12 as a criterion to identify the block-funded articles to be matched with the grant-funded articles and rely on CEM to obtain a better balance among covariates than the PSM. With the CEM procedure, we coarsen the support of the joint distribution of the covariates into a set of strata, then allocate each article to a unique stratum. We drop strata that do not contain at least one grant-funded article and one block-funded article. We end up with 6,229 strata. Finally, we match each grant-funded article with the block-funded article allocated in the same stratum. If a stratum contains multiple articles, we match a grant-funded article with the most similar block-funded article within the same stratum relying on the Propensity Score Matching as proposed in the main analysis. The difference from the PSM procedure of our main analysis is that with CEM the articles are stratified ex-ante using a nonparametric approach. This allows us to guarantee the covariate balance ex-ante. The cost is given by the loss of 212 grant-funded unmatched publications. We end up with a matched sample composed of 6,229 publications supported by competitive grants paired with 6,229 publications supported by block funding. Appendix U explains the CEM matching procedure in detail and reports the covariate balance table pre- and post-matching (Table U1). With the CEM, after matching all the articles' and authors' characteristics are perfectly balanced between the grant-funded and block-funded groups of publications. All the covariate p-values equal 1.000.

Table 21 reports the difference in the funding effects on the articles' impact in the short and long run. In the long run, the result is consistent with that reported in Table 15, evidencing how publications supported by competitive grants are more impactful than those supported by block funding. The magnitude slightly decreases, moving from a grant-funded relative effect in the long run of +6.93% in Table 15, to +5.95% in Table 21. Instead, the difference in the funding effect in the short run is now statistically significant. The CEM procedure shows that grant-funded publications are less impactful than block-funded publications in the short run. The grant-funded relative effect is negative and equals -4.39% (Table 21) in the short run. In our main analysis (Table 15), we did not find a statistically significant difference in the short run, although the direction of the effect was the same: grant-funded articles showed -3.4% fewer citations than block-funded articles.

Table 21. Funding effects on publications' number of citations received.

6,229 Grant-funded + 6,229 Block-funded	(1) **Short run** t to t+2, citations	(2) **Long run** t+3 to t+5, citations
Average citations Grant-funded articles (A)	6.365	8.887
Average citations Block-funded articles (B)	6.658	8.387
Grant-funded effect (A-B)	-0.292 *	0.499 **
Grant-funded relative effect (A-B)/B	-4.39% *	+5.95% **
t-statistic	*-1.826*	*2.072*
p-value	*0.068*	*0.038*

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

3.7 Conclusion

Europe is relying more and more on competitive models to allocate research funding. Over the last fifteen years, the dominant block funding model has been increasingly replaced by the competitive grant model. This trend is probably due to the policymakers' belief that funding science through a competitive grant model is more effective than relying on a block funding model, leading to more impactful research. National funding agencies have recently sprung up in several European countries. In France, *l'Agence Nationale de la recherche* (ANR) was funded in 2005 to support French research excellence through a competitive allocation of funds to French scientists.

In this study, using all the articles resulting from grants distributed by the ANR agency between 2009 and 2013, we compare the impact of scientific articles supported by competitive grants with the impact of articles resulting from block funding, both in the short and long run. Moreover, we investigate whether the funding effects differ across four fields of research. We rely on publications' acknowledgment data to identify publications supported by the two different funding models. Using a comprehensive set of articles' and authors' characteristics, we propose a Propensity Score Matching approach to assess the effects of funding.

We find that articles resulting from competitive grants receive significantly more citations than articles resulting from block funding. This result is in line with the literature supporting the beneficial effects of the competitive grant model, which facilitates successful research beyond institutional block funding (Heyard and Hottenrott, 2021). Interestingly, in our analysis, we find that grant-funded articles are more impactful than block-funded articles in the long run (+6.93%), but not in the short run. Wang et al. (2017) report strong evidence that breakthrough research takes more time to collect citations. They show that highly novel articles and breakthrough research share similar citation patterns receiving more citations in the long run according to the "high risk/high gain" profile. Specifically, they found that highly novel articles outpace non-novel articles in terms of citation impact three years after the publication date. Coherently with Wang et al. (2017), in our analysis, we find that grant-funded publications are more impactful than block-funded publications in the long run (i.e., from three to five years after publication). Following Wang et al. (2017), we can interpret our results as the ANR funding agency's effort to support breakthrough research.

The higher impact of grant-funded articles might also derive from the additional resources provided by the competitive grants. State-of-the-art equipment, access to data, additional workforce in the lab, and additional financial resources in support of the result dissemination might be ensured through the grant and lead to more impactful research. Another possible interpretation of our findings concerns the cumulative citation process. Researchers awarded a competitive grant are more likely to succeed again in obtaining further grants in the future (Bol et al., 2018). Research projects led by the same researcher are likely to be based on similar topics, increasing the probability of self-citation of the researcher's work. Following this line of reasoning, an increase in citations tends to appear in the long run.

Our paper highlights other interesting results. We find that funding effects differ by field of research. Publications in Life sciences and Medicine, and Engineering, are those showing the highest increase of citations in the long run when supported by competitive grants (+15.06% and +13.87%, respectively). Engineering is the only field of research where publications are more impactful when supported by competitive grants also in the short run (+8.15% in the short run). This result is in line with previous studies that evidence a beneficial effect of grant funding on research outcomes in Engineering and Nanotechnology (Hottenrott and Thorwarth, 2011; Beaudry and Allaoui, 2012; Wang and Shapira, 2015; Tahmooresnejad and Beaudry, 2019). The finding suggests that high-quality research in Engineering is grant funding dependent, probably due to the additional resources from the grants that allow researchers in engineering to buy state-of-the-art equipment needed to carry out impactful research. Interestingly, Mathematics is the only field of research where publications do not benefit from competitive grants. On the contrary, publications resulting from grant funding are less impactful than block-funded publications in the short run (-13.08%). This different finding for Mathematics can be explained by the peculiarities of this field: mathematicians often work alone, and they have less need to go to the lab and collaborate with other researchers, compared to the other fields. Moreover, the equipment necessary to conduct research in Mathematics is often minimal. Thus, they do not necessarily need additional resources from grants to produce impactful research.

We also show in a robustness check that when including articles supported by multiple grants among the sample of grant-funded articles (and not only articles supported exclusively by ANR grants), the impact of the competitive grants relative to block funding is higher (+27.23% in the short run and +25.56% in the long run). This finding supports the idea of a beneficial effect of competitive grants in science and goes in the opposite direction to what Mali et al. (2017) found. They stated that public grants support impactful research only if researchers' funding comes from one single source.

Our study is not exempt from limitations. First, we are limited to using articles published from 2009 to 2013, due to WOS and MAG data accuracy. According to this time window, we can show the effect of funding on articles' citations only up to 6 years after the publication date. With a longer time window, we could have been able to interpret our results better. Second, when matching articles, we cannot control the manuscripts' content and other unobserved factors influencing the selection process made by the ANR review board.

3.8 Appendix of Chapter 3

Appendix O

This appendix reports the tables of the average marginal effects of the articles' probability of being supported by an ANR grant, calculated from the respective logit coefficients, and the covariate balance tables pre- and post-matching for each of the four fields of research we consider. For each field, we run a separate *nearest neighbor* Propensity Score Matching to pair the grant-funded articles with the block-funded articles that have the highest propensity score similarity and belong to the same research field.

We assign an article to a research field when it is published in a journal classified at least in that research field. Journals might be classified in multiple research fields, thus, we allow for double-counting of articles. When running the Propensity Score Matching by research field, we include the research fields that are different than the focal one among the covariates used for the matching. In so doing, we ensure that each grant-funded article published in a journal classified in more than one field is paired with a block-funded article having the same multiple classifications, i.e., published in a similar journal.

Tables O1 and O2 refer to articles in Life sciences and Medicine. We identify 1,708 grant-funded articles in Life sciences and Medicine matched with 1,708 similar block-funded articles drawn from a pool of 37,657 articles in Life sciences and Medicine. Table O1 reports the average marginal effects calculated from the estimated logit coefficients that predict the probability that an article classified in Life sciences and Medicine is supported by a grant. Differently from the main analysis of Table 13, articles in Life sciences and Medicine are more likely to result from grants when they have between 2 and 4 authors. Moreover, having a female author is positively related to the likelihood of being supported by grants. Interestingly, when the article is published in a journal classified in an additional field of research besides Life sciences and Medicine, i.e., a multidisciplinary journal, the likelihood that ANR grants support the paper increases. Table O2 shows how after the Propensity Score Matching, all the articles' and authors' characteristics of the group of grant-funded articles in Life sciences and Medicine are statistically equivalent to those of the group of block-funded articles, at standard significance levels.

Tables O3 and O4 refer to articles in Mathematics. We match 1,806 grant-funded articles in Mathematics with 1,806 similar block-funded articles drawn from a pool of 14,453 articles in Mathematics. Table O3 shows the average marginal effects of the article's probability of being supported by a grant, calculated from the estimated logit coefficients. Differently from the main analysis of Table 13, the presence of an international author further decreases the likelihood that articles in Mathematics are supported by grants (the average marginal effect equals -7.3% in the main analysis, while it equals -11% when considering articles in Mathematics). Furthermore, articles in Mathematics having a number of backward citations belonging to the third quartile of the backward citation distribution are more likely to result from ANR grants. Contrary to articles in Life sciences and Medicine, articles in Mathematics published in multidisciplinary journals do not benefit from an increased probability of being supported by ANR grants (when published in journals classified both in Mathematics and Engineering, the probability decreases by 6%). Table O4 shows that after the Propensity Score Matching, all the articles' and authors' characteristics of the group of grant-funded publications in Mathematics are statistically equivalent to those of the group of block-funded articles, at standard significance levels.

Tables O5 and O6 refer to articles in Engineering. We identify 1,489 grant-funded articles in Engineering that we match with 1,489 similar block-funded articles drawn from a pool of 15,261 articles in Engineering. Table O5 shows the average marginal effects of the article's probability of being supported by a grant, calculated from the estimated logit coefficients. Compared to the main analysis of Table 13, the presence of more than 4 authors further increases the likelihood that articles in Engineering are supported by grants (from +5.8% in Table 13 to +9.2% in Table O5). Similarly to articles in Mathematics, the presence of an international author further decreases the likelihood that articles in Engineering are supported by grants (-10%, compared to -7.3% of the main analysis of Table 13). Contrary to the articles in Mathematics but similarly to articles in Life sciences and Medicine, articles in Engineering are more likely to result from competitive grants when published in multidisciplinary journals. Table O6 shows that after the Propensity Score Matching, all the articles' and authors' characteristics are statistically equivalent between the group of grant-funded publications and the group of block-funded articles, at standard significance levels.

Finally, tables O7 and O8 refer to articles in Physical sciences. We match 3,719 grant-funded articles in Physical sciences with 3,719 similar block-funded articles drawn from a pool of 29,081 articles in Physical sciences. Table O7 shows the average marginal effects of the article's probability of being supported by a grant, calculated from the estimated logit coefficients. Compared to the main analysis of Table 13, the presence of multiple authors further increases the likelihood that articles in Physical sciences are supported by grants. Articles having from 2 to 4 authors have an 8.6% more probability of being supported by competitive grants than articles with a single author (compared to +4.4% of the main analysis of Table 13). For articles with more than 4 authors, the probability is 14% higher than in single-author articles (compared to +5.8% in Table 13). Similarly to articles in Mathematics and Engineering, the presence of an international author further decreases the probability that the article is supported by grants (-12%, compared to -7.3% in Table 13). Finally, contrary to all the other fields of research, articles in Physical sciences reporting a single affiliation do not have a different probability of being supported by competitive grants than articles reporting multiple authors' affiliations. Table O8 shows that after the Propensity Score Matching, all the articles' and authors' characteristics are statistically equivalent between the group of grant-funded publications and the group of block-funded publications, at standard significance levels.

The equivalence of all the article's and authors' characteristics between grant-funded publications and block-funded publications, after the Propensity Score Matching, for all the four research fields analyzed (Tables O2, O4, O6, and O8) supports the reliability of our results across research fields presented in Table 16.

Life sciences and Medicine

Table O1. Average marginal effects of the article's probability of being supported by an ANR grant for articles in Life sciences and Medicine.

	(1) **Grant-funded (=1) versus Block-funded (=0)**
Single-author article	Ref.
Multi-author article (from 2 to 4 authors)	0.042*** (0.0068)
Multi-author article (more than 4 authors)	0.039*** (0.0070)
At least one international author	-0.037*** (0.0029)
At least one female author	0.0067** (0.0028)
At least one top-affiliate author	0.012*** (0.0021)
Multiple affiliations	0.011*** (0.0026)
Backward citations Q1	Ref.
Backward citations Q2	0.033*** (0.0048)
Backward citations Q3	0.059*** (0.0045)
Backward citations Q4	0.085*** (0.0044)
Life sciences and Medicine	Ref.
Mathematics	0.051*** (0.0045)
Engineering	0.029*** (0.0036)
Physical sciences	0.034*** (0.0025)
Year of publication	0.0073*** (0.00071)
Pseudo R2	0.1255
Number of articles	39,365

NOTE: Average marginal effects are calculated from estimated logit coefficients. Standard errors in parentheses. Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Table O2. Means of the articles' and authors' observable characteristics for grant- and block-funded articles in Life sciences and Medicine, before (Columns 1 and 2) and after (Columns 3 and 4) the Propensity Score Matching.

	Sample before PSM			PSM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	1,708	37,657		1,708	1,708	
Covariates						
Single-author article	0.023***	0.076	0.000	0.023	0.022	0.817
Multi-author article (from 2 to 4 authors)	0.403***	0.343	0.000	0.403	0.407	0.780
Multi-author article (more than 4 authors)	0.574	0.581	0.565	0.574	0.571	0.836
At least one international author	0.153***	0.239	0.000	0.153	0.159	0.637
At least one female author	0.828***	0.776	0.000	0.828	0.827	0.964
At least one top-affiliate author	0.391***	0.303	0.000	0.391	0.391	0.972
Multiple affiliations	0.783***	0.711	0.000	0.783	0.790	0.616
Backward citations Q1	0.059***	0.280	0.000	0.059	0.054	0.505
Backward citations Q2	0.134***	0.233	0.000	0.134	0.131	0.762
Backward citations Q3	0.271***	0.240	0.004	0.271	0.270	0.969
Backward citations Q4	0.536***	0.247	0.000	0.536	0.545	0.583
Life sciences and Medicine	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Mathematics	0.066***	0.017	0.000	0.066	0.055	0.196
Engineering	0.111***	0.033	0.000	0.111	0.110	0.956
Physical sciences	0.287***	0.110	0.000	0.287	0.292	0.734
Year of publication	2011.34***	2010.94	0.000	2011.34	2011.31	0.604

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Mathematics

Table O3. Average marginal effects of the article's probability of being supported by an ANR grant for articles in Mathematics.

	(1) **Grant-funded (=1) versus Block-funded (=0)**
Single-author article	Ref.
Multi-author article (from 2 to 4 authors)	0.049*** (0.0073)
Multi-author article (more than 4 authors)	0.068*** (0.012)
At least one international author	-0.11*** (0.0070)
At least one female author	-0.014*** (0.0053)
At least one top-affiliate author	0.018*** (0.0050)
Multiple affiliations	0.021*** (0.0061)
Backward citations Q1	Ref.

Backward citations Q2	0.028*** (0.0067)
Backward citations Q3	0.040*** (0.0068)
Backward citations Q4	0.038*** (0.0076)
Mathematics	Ref.
Life sciences and Medicine	0.011 (0.011)
Engineering	-0.060*** (0.0065)
Physical sciences	-0.0018 (0.0063)
Year of publication	0.023*** (0.0017)
Pseudo R2	0.056
Number of articles	16,259

NOTE: Average marginal effects are calculated from estimated logit coefficients. Standard errors in parentheses. Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Table O4. Means of the articles' and authors' observable characteristics for grant- and block-funded articles in Mathematics, before (Columns 1 and 2) and after (Columns 3 and 4) the Propensity Score Matching.

	Sample before PSM			PSM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	1,806	14,453		1,806	1,806	
Covariates						
Single-author article	0.185***	0.232	0.000	0.185	0.190	0.702
Multi-author article (from 2 to 4 authors)	0.729***	0.691	0.001	0.729	0.736	0.625
Multi-author article (more than 4 authors)	0.085	0.078	0.267	0.085	0.073	0.175
At least one international author	0.159***	0.305	0.000	0.159	0.162	0.821
At least one female author	0.358	0.362	0.785	0.358	0.350	0.602
At least one top-affiliate author	0.417***	0.356	0.000	0.417	0.421	0.840
Multiple affiliations	0.625	0.609	0.182	0.625	0.628	0.863
Backward citations Q1	0.223***	0.306	0.000	0.223	0.215	0.546
Backward citations Q2	0.297	0.295	0.820	0.297	0.306	0.587
Backward citations Q3	0.285***	0.238	0.000	0.285	0.291	0.659
Backward citations Q4	0.195***	0.161	0.001	0.195	0.188	0.612
Life sciences and Medicine	0.062***	0.044	0.002	0.062	0.052	0.197
Mathematics	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Engineering	0.169***	0.249	0.000	0.169	0.159	0.419
Physical sciences	0.189	0.179	0.288	0.189	0.188	0.898
Year of publication	2011.41***	2010.93	0.000	2011.41	2011.40	0.831

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Engineering

Table O5. Average marginal effects of the article's probability of being supported by an ANR grant for articles in Engineering.

	(1) **Grant-funded (=1) versus Block-funded (=0)**
Single-author article	Ref.
Multi-author article (from 2 to 4 authors)	0.043*** (0.011)
Multi-author article (more than 4 authors)	0.092*** (0.012)
At least one international author	-0.10*** (0.0068)
At least one female author	-0.0056 (0.0047)
At least one top-affiliate author	0.026*** (0.0048)
Multiple affiliations	0.0085* (0.0051)
Backward citations Q1	Ref.
Backward citations Q2	0.033*** (0.0066)
Backward citations Q3	0.042*** (0.0066)
Backward citations Q4	0.058*** (0.0067)
Engineering	Ref.
Life sciences and Medicine	0.025*** (0.0068)
Mathematics	0.018*** (0.0060)
Physical sciences	0.038*** (0.0054)
Year of publication	0.018*** (0.0015)
Pseudo R2	0.0817
Number of articles	16,750

NOTE: Average marginal effects are calculated from estimated logit coefficients. Standard errors in parentheses. Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Table O6. Means of the articles' and authors' observable characteristics for grant- and block-funded articles in Engineering, before (Columns 1 and 2) and after (Columns 3 and 4) the Propensity Score Matching.

	Sample before PSM			PSM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	1,489	15,261		1,489	1,489	
Covariates						
Single-author article	0.038***	0.070	0.000	0.038	0.036	0.770
Multi-author article (from 2 to 4 authors)	0.510***	0.632	0.000	0.510	0.516	0.742
Multi-author article (more than 4 authors)	0.453***	0.298	0.000	0.453	0.449	0.825
At least one international author	0.118***	0.292	0.000	0.118	0.118	0.955
At least one female author	0.596***	0.527	0.000	0.596	0.602	0.737
At least one top-affiliate author	0.298***	0.234	0.000	0.298	0.302	0.810
Multiple affiliations	0.680	0.666	0.273	0.680	0.685	0.753
Backward citations Q1	0.159***	0.271	0.000	0.159	0.163	0.765
Backward citations Q2	0.269	0.285	0.183	0.269	0.271	0.934
Backward citations Q3	0.291***	0.252	0.001	0.291	0.295	0.809
Backward citations Q4	0.281***	0.192	0.000	0.281	0.271	0.566
Life sciences and Medicine	0.127***	0.081	0.000	0.127	0.116	0.370
Mathematics	0.205***	0.236	0.005	0.205	0.197	0.615
Engineering	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Physical sciences	0.727***	0.609	0.000	0.727	0.733	0.710
Year of publication	2011.42***	2010.91	0.000	2011.42	2011.45	0.550

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Physical sciences

Table O7. Average marginal effects of the article's probability of being supported by an ANR grant for articles in Physical sciences.

	(1) **Grant-funded (=1) versus Block-funded (=0)**
Single-author article	Ref.
Multi-author article (from 2 to 4 authors)	0.086*** (0.0083)
Multi-author article (more than 4 authors)	0.14*** (0.0090)
At least one international author	-0.12*** (0.0050)
At least one female author	-0.0095** (0.0039)
At least one top-affiliate author	0.035*** (0.0036)
Multiple affiliations	0.0065 (0.0041)
Backward citations Q1	Ref.

Backward citations Q2	0.032*** (0.0058)
Backward citations Q3	0.047*** (0.0056)
Backward citations Q4	0.058*** (0.0055)
Physical sciences	Ref.
Life sciences and Medicine	-0.024*** (0.0052)
Mathematics	0.012* (0.0061)
Engineering	-0.016*** (0.0038)
Year of publication	0.019*** (0.0012)
Pseudo R2	0.0665
Number of articles	32,800

NOTE: Average marginal effects are calculated from estimated logit coefficients. Standard errors in parentheses. Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Table O8. Means of the articles' and authors' observable characteristics for grant- and block-funded articles in Physical sciences, before (Columns 1 and 2) and after (Columns 3 and 4) the Propensity Score Matching.

	Sample before PSM			PSM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	3,719	29,081		3,719	3,719	
Covariates						
Single-author article	0.045***	0.093	0.000	0.045	0.045	0.956
Multi-author article (from 2 to 4 authors)	0.506***	0.559	0.000	0.506	0.508	0.799
Multi-author article (more than 4 authors)	0.449***	0.348	0.000	0.449	0.446	0.816
At least one international author	0.144***	0.321	0.000	0.144	0.144	0.947
At least one female author	0.619***	0.580	0.000	0.619	0.621	0.848
At least one top-affiliate author	0.405***	0.320	0.000	0.405	0.405	0.981
Multiple affiliations	0.685	0.681	0.571	0.685	0.689	0.708
Backward citations Q1	0.130***	0.206	0.000	0.130	0.131	0.863
Backward citations Q2	0.222***	0.241	0.009	0.222	0.220	0.867
Backward citations Q3	0.283***	0.253	0.000	0.283	0.286	0.777
Backward citations Q4	0.366***	0.301	0.000	0.366	0.363	0.810
Life sciences and Medicine	0.132*	0.143	0.062	0.132	0.124	0.297
Mathematics	0.092	0.089	0.550	0.092	0.088	0.543
Engineering	0.291***	0.320	0.000	0.291	0.286	0.645
Physical sciences	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Year of publication	2011.26***	2010.82	0.000	2011.26	2011.25	0.831

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Appendix P

This appendix disentangles the difference in the funding effects on the articles' number of citations received in each of the 6 years after publication.

Table P1 reports the results. In the first year after publication, articles resulting from competitive grants receive 30.85% fewer citations than articles resulting from block funding. In the second year after publication, there is no statistically significant difference between grant-funded and block-funded articles' impact. From the third year after publication until the sixth year, grant-funded articles receive more citations than block-funded articles. The largest impact is in the fifth year, when grant-funded articles receive 8.57% more citations than block-funded articles.

Table P1. Funding effects on publications' number of citations received each year.

6,441 Grant-funded + 6,441 Block-funded	(1) **First year** citations	(2) **Second year** citations	(3) **Third year** citations	(4) **Fourth year** citations	(5) **Fifth year** citations	(6) **Sixth year** citations
Avg. citations Grant-funded articles (A)	0.778	2.473	3.151	3.231	2.992	2.720
Avg. citations Block-funded articles (B)	1.126	2.544	2.958	3.039	2.756	2.568
Grant-funded effect (A-B)	-0.347 ***	-0.071	0.193 **	0.192 **	0.236 ***	0.152 *
Grant-funded relative effect (A-B)/B	-30.85% ***	-2.79%	+6.52% **	+6.32% **	+8.57% ***	+5.90% *
t-statistic	*-10.973*	*-1.082*	*2.514*	*2.289*	*2.832*	*1.816*
p-value	*0.000*	*0.279*	*0.012*	*0.022*	*0.005*	*0.069*

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Appendix Q

This appendix relies on SCOPUS citation data to compare to effects of funding on publications' impact. The methodology to compare articles remains the same as described in section 3.3. Different from the main analysis, we use SCOPUS average yearly citations instead of MAG annual citations to assess the articles' impact. Specifically, we retrieved the cumulated number of each article's citations in 2018 from SCOPUS. Then, we calculate the number of average yearly citations received by each article by dividing the cumulated number of citations by the years elapsed between the article's publication date and 2018.

The advantage of using SCOPUS average yearly citations is that we allow for a time window longer than 6 years for articles published before 2013[87]. The disadvantages concern the use of different time windows for different article cohorts and the impossibility of distinguishing between the short and long run.

Table Q1 reports the results. Articles resulting from competitive grants receive, on average, 4.40% more yearly citations than articles resulting from block funding.

[87] The longest time window regards articles published in 2009 and equals 10 years.

Table Q1. Funding effects on publications' average number of yearly citations received.

	(1)
	From publication date to 2018
6,441 Grant-funded + 6,441 Block-funded	Avg. yearly citations
Average yearly citations Grant-funded articles (A)	2.983
Average yearly citations Block-funded articles (B)	2.857
Grant-funded effect (A-B)	0.126 *
Grant-funded relative effect (A-B)/B	+4.40% *
t-statistic	*1.718*
p-value	*0.086*

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1. Citations are calculated using SCOPUS data.

Appendix R

This appendix reports the table of the average marginal effects of the articles' probability of being supported by an ANR grant (Table R1) and the covariate balance table pre- and post-matching (Table R2) for the robustness check of section 3.6.3. In this robustness check, we add to our main model three covariates describing the academic status of the French researchers who author the articles of our sample. Specifically, we include the dummy variable *At least a French star author* to identify the presence of at least one French star scientist among the article's authors, the dummy variable *At least a French senior author* that identifies the presence of at least one French senior scientist among the article's authors, and the dummy variable *At least a French Ph.D. student author* that identifies the presence of at least one French Ph.D. student among the article's authors.

Table R1 shows that when a French star scientist is among the article's authors the probability that the article is supported by ANR grant increases by 2%, while the presence of a Ph.D. student among the article's authors increases the same probability by 1.3%. The presence of a French senior scientist among the authors does not statistically influence the likelihood that the article results from ANR grants. The inclusion of these three new variables does not qualitatively change the average marginal effects of the other estimated coefficients.

Considering three new covariates makes it more difficult to maintain grant-funded articles and their potential matchable block-funded articles on the common support, i.e., to ensure that grant-funded articles and block-funded articles have a similar propensity score distribution after matching. To force it, we include the Caliper in the *nearest neighbor* Propensity Score Matching. The Caliper measures the number of standard deviations of the propensity score distance allowed between a grant-funded article and its potential matchable block-funded article. We impose a Caliper of 0.000001. If no matches are available within the Caliper, the articles are discarded. Doing so, we are willing to lose articles to make sure we can balance the covariates of the grant-funded and block-funded articles after the matching. Table R2 reports the covariate balance table before and after the PSM. After the matching, we remain with 5,537 grant-funded articles matched with 5,537 similar block-funded articles. All the articles' and authors' characteristics are statistically equivalent between the group of grant-funded articles and that of block-funded articles, at standard significance levels. 53.5% of both grant-funded and block-funded publications have *At least a French star author*, 91.8% have *At least a French senior author*, and 38% have *At least a French Ph.D. student author*, after the Propensity Score Matching.

Table R1. Average marginal effects of the article's probability of being supported by an ANR grant.

	Grant-funded (=1) versus Block-funded (=0)
Single-author article	Ref.
Multi-author article (from 2 to 4 authors)	0.037*** (0.0039)
Multi-author article (more than 4 authors)	0.043*** (0.0045)
At least one international author	-0.068*** (0.0027)
At least one female author	-0.0078*** (0.0021)
At least one top-affiliate author	0.018*** (0.0019)
Multiple affiliations	0.011*** (0.0022)
Backward citations Q1	Ref.
Backward citations Q2	0.025*** (0.0030)
Backward citations Q3	0.040*** (0.0029)
Backward citations Q4	0.057*** (0.0028)
Life sciences and Medicine	Ref.
Mathematics	0.075*** (0.0025)
Engineering	-0.0045** (0.0022)
Physical sciences	0.069*** (0.0020)
Year of publication	0.014*** (0.00064)
At least a French star author	**0.020*** (0.0021)**
At least a French senior author	**-0.0014 (0.0032)**
At least a French Ph.D. student author	**0.013*** (0.0020)**
Pseudo R2	0.0892
Number of articles	83,056

NOTE: Average marginal effects are calculated from estimated logit coefficients. Standard errors in parentheses. Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Table R2. Means of the articles' and authors' observable characteristics for grant- and block-funded publications, before (Columns 1 and 2) and after (Columns 3 and 4) the Propensity Score Matching.

	Sample before PSM			PSM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	6,441	76,615		5,537	5,537	
Covariates						
Single-author article	0.075***	0.109	0.000	0.075	0.075	1.000
Multi-author article (from 2 to 4 authors)	0.539***	0.483	0.000	0.537	0.537	0.985
Multi-author article (more than 4 authors)	0.386***	0.408	0.001	0.387	0.387	0.984
At least one international author	0.153***	0.279	0.000	0.153	0.153	1.000
At least one female author	0.596***	0.628	0.000	0.613	0.613	1.000
At least one top-affiliate author	0.399***	0.315	0.000	0.384	0.384	0.984
Multiple affiliations	0.693*	0.682	0.058	0.706	0.706	0.983
Backward citations Q1	0.145***	0.269	0.000	0.142	0.142	1.000
Backward citations Q2	0.225***	0.249	0.000	0.223	0.223	0.982
Backward citations Q3	0.278***	0.241	0.000	0.281	0.281	0.983
Backward citations Q4	0.352***	0.240	0.000	0.354	0.354	1.000
Life sciences and Medicine	0.265***	0.492	0.000	0.259	0.259	1.000
Mathematics	0.280***	0.189	0.000	0.250	0.250	0.982
Engineering	0.231***	0.199	0.000	0.199	0.199	1.000
Physical sciences	0.577***	0.380	0.000	0.580	0.580	0.985
Year of publication	2011.31***	2010.90	0.000	2011.29	2011.29	0.989
At least a French star author	**0.534*** **	**0.470**	**0.000**	**0.535**	**0.535**	**0.985**
At least a French senior author	**0.899*** **	**0.867**	**0.000**	**0.918**	**0.918**	**1.000**
At least a French Ph.D. student author	**0.400*** **	**0.326**	**0.000**	**0.380**	**0.381**	**0.984**

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Appendix S

This appendix reports the covariate balance table pre- and post-matching for the robustness check of section 3.6.4. Specifically, Table S1 refers to the Propensity Score Matching conditional on an exact matching on the articles' *Year of publication*, while Table S2 refers to the Propensity Score Matching conditional on an exact matching both on the articles' *Year of publication* and *Journal of publication*.

Table S1 shows that, after the matching with the exact condition on the publication year, all the articles' and authors' characteristics are statistically equivalent between the group of grant-funded articles and that of block-funded articles, at standard significance levels. The *Year of publication* is perfectly balanced between grant-funded articles and block-funded articles due to the exact condition imposed. Its mean equals 2011.314 for both groups.

Table S2 shows that, including a double exact condition, i.e., both on the *Year of publication* and *Journal of publication*, we are first forced to lose observations. We remain with 1,643 grant-funded articles matched with 1,643 similar block-funded articles. This reduction is due to the fact that for many grant-funded articles we do not find block-funded articles that are published on the same journal and in the same year. Moreover, with the aim of keeping the balance between the covariates, we introduce the Caliper at 0.02. Doing so, we match a grant-funded article with a block-funded article only when the latter is published on the same journal, in the same year, and has a similarity distance within 0.02 standard deviations of the propensity score of the former.

Table S2 shows that, post matching, all the articles' and authors' characteristics are statistically equivalent between the group of grant-funded articles and that of block-funded articles, at standard significance levels. The *Year of publication* is perfectly balanced between grant-funded articles and block-funded articles with a mean equal to 2011.133 for both groups. Table S2 shows also that the sample after matching, composed of 1,643 grant-funded articles and 1,643 block-funded articles, includes 71% of articles published in journals classified at least in Physical sciences.

Table S1. Means of the articles' and authors' observable characteristics for grant- and block-funded publications, before (Columns 1 and 2) and after (Columns 4 and 5) the Propensity Score Matching conditional on an exact matching on the year of publication.

	Sample before PSM			PSM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	6,441	76,615		6,441	6,441	
Covariates						
Single-author article	0.075***	0.109	0.000	0.075	0.076	0.815
Multi-author article (from 2 to 4 authors)	0.539***	0.483	0.000	0.539	0.546	0.437
Multi-author article (more than 4 authors)	0.386***	0.408	0.001	0.386	0.378	0.355
At least one international author	0.153***	0.279	0.000	0.153	0.156	0.643
At least one female author	0.596***	0.628	0.000	0.596	0.595	0.872
At least one top-affiliate author	0.399***	0.315	0.000	0.399	0.398	0.900
Multiple affiliations	0.693*	0.682	0.058	0.693	0.693	1.000
Backward citations Q1	0.145***	0.269	0.000	0.145	0.143	0.802
Backward citations Q2	0.225***	0.249	0.000	0.225	0.225	0.950
Backward citations Q3	0.278***	0.241	0.000	0.278	0.282	0.624
Backward citations Q4	0.352***	0.240	0.000	0.352	0.350	0.825
Life sciences and Medicine	0.265***	0.492	0.000	0.265	0.255	0.178
Mathematics	0.280***	0.189	0.000	0.280	0.277	0.709
Engineering	0.231***	0.199	0.000	0.231	0.228	0.660
Physical sciences	0.577***	0.380	0.000	0.577	0.578	0.957
Year of publication	**2011.31*** **	**2010.90**	**0.000**	**2011.314**	**2011.314**	**1.000**

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Table S2. Means of the articles' and authors' observable characteristics for grant- and block-funded publications, before (Columns 1 and 2) and after (Columns 3 and 4) the Propensity Score Matching conditional on an exact matching on the year and journal of publication.

	Sample before PSM			PSM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	6,441	76,615		1,643	1,643	
Covariates						
Single-author article	0.075***	0.109	0.000	0.077	0.091	0.147
Multi-author article (from 2 to 4 authors)	0.539***	0.483	0.000	0.492	0.500	0.650
Multi-author article (more than 4 authors)	0.386***	0.408	0.001	0.432	0.410	0.203
At least one international author	0.153***	0.279	0.000	0.160	0.167	0.572
At least one female author	0.596***	0.628	0.000	0.629	0.605	0.151
At least one top-affiliate author	0.399***	0.315	0.000	0.413	0.406	0.696
Multiple affiliations	0.693*	0.682	0.058	0.704	0.696	0.621
Backward citations Q1	0.145***	0.269	0.000	0.168	0.165	0.815
Backward citations Q2	0.225***	0.249	0.000	0.214	0.205	0.548

Backward citations Q3	0.278***	0.241	0.000	0.276	0.264	0.432
Backward citations Q4	0.352***	0.240	0.000	0.342	0.366	0.155
Life sciences and Medicine	0.265***	0.492	0.000	0.181	0.180	0.928
Mathematics	0.280***	0.189	0.000	0.211	0.211	1.000
Engineering	0.231***	0.199	0.000	0.187	0.186	0.893
Physical sciences	0.577***	0.380	0.000	0.708	0.707	0.939
Year of publication	**2011.31*** **	**2010.90**	**0.000**	**2011.133**	**2011.133**	**1.000**

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Appendix T

This appendix reports the table of the average marginal effects of the articles' probability of being supported by any type of competitive grant (with at least an ANR grant) (Table T1) and the covariate balance table pre- and post-matching (Table T2) for the robustness check of section 3.6.5. In this robustness check, we extend the group of grant-funded publications by considering the articles resulting from multiple competitive grants, conditional on having at least one ANR grant. Doing so, we allow for the presence of other non-ANR sources of funding along with ANR grants.

Table T1 reports the average marginal effects of the likelihood that an article is supported by any type of competitive grant (with at least an ANR grant) calculated on a sample of 100,565 publications, divided into 23,950 grant-funded articles and 76,615 block-funded articles. The number of block-funded articles remains the same as in our main analysis, while the sample of grant-funded articles increases from 6,441 to 23,950 articles due to the inclusion of publications resulting from multiple grants. Compared to the main analysis that includes only grant-funded articles supported solely by ANR grants, Table T1 shows that the presence of multiple authors further increases the probability that an article is supported by any type of competitive grant. The presence of 2 to 4 authors enhances this probability by 14% compared to single-author articles, while the presence of more than 4 authors increases the probability by 21%. This finding is explained by the fact that we are considering also competitive grants awarded by sources different than ANR, thus, the higher the number of authors the greater the probability that one of them results from any type of competitive grant (conditional on the fact that one of them is supported by ANR grant). For the same reason, we see in Table T1 that the presence of an international author is now associated with an increase in the probability that an article is supported by competitive grants (+1%, compared to -7.3% of the main analysis of Table 13). We also find that the presence of a female author is now positively associated with the probability that an article is supported by competitive grants (+1.1%, compared to -0.71% of Table 13), and the presence of at least one author affiliated with a top institution and a number of backward citations belonging to the last quartile of the backward citation distribution are now more relevant (from +2% of Table 13 to +7.6% of Table T1, and from +6% of Table 13 to +25% of Table T1, respectively).

Table T2 reports the covariate balance table before and after the Propensity Score Matching. To maintain grant-funded and block-funded publications on the common support, i.e., to ensure that grant-funded articles and block-funded articles have a similar propensity score distribution after matching, we introduce the Caliper at 0.0001. Doing so, we discard the matches where the distance between a grant-funded article and a block-funded article is greater than 0.0001 standard deviations of the propensity score distance. This procedure forces us to lose observations. After the matching, we end up with 21,381 grant-funded articles matched with 21,381 similar block-funded articles. Table T2 shows that, after the matching, all the articles' and authors' characteristics are statistically equivalent between the group of grant-funded articles and that of block-funded articles, at standard significance levels.

Table T1. Average marginal effects of the article's probability of being supported any type of competitive grant (with at least an ANR grant).

	(1) **Grant-funded (=1) versus Block-funded (=0)**
Single-author article	Ref.
Multi-author article (from 2 to 4 authors)	0.14*** (0.0067)
Multi-author article (more than 4 authors)	0.21*** (0.0071)
At least one international author	0.010*** (0.0029)
At least one female author	0.011*** (0.0031)
At least one top-affiliate author	0.076*** (0.0026)
Multiple affiliations	0.026*** (0.0036)
Backward citations Q1	Ref.
Backward citations Q2	0.091*** (0.0046)
Backward citations Q3	0.16*** (0.0043)
Backward citations Q4	0.25*** (0.0040)
Life sciences and Medicine	Ref.
Mathematics	0.10*** (0.0037)
Engineering	-0.052*** (0.0035)
Physical sciences	0.099*** (0.0027)
Year of publication	0.033*** (0.00088)
Pseudo R2	0.1144
Number of articles	100,565

NOTE: Average marginal effects are calculated from estimated logit coefficients. Standard errors in parentheses. Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Table T2. Means of the articles' and authors' observable characteristics for grant- and block-funded publications considering multiple competitive grants, before (Columns 1 and 2) and after (Columns 3 and 4) the Propensity Score Matching.

	Sample before PSM			PSM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	23,950	76,615		21,381	21,381	
Covariates						
Single-author article	0.034***	0.109	0.000	0.036	0.037	0.625
Multi-author article (from 2 to 4 authors)	0.424***	0.483	0.000	0.460	0.461	0.869
Multi-author article (more than 4 authors)	0.542***	0.408	0.000	0.504	0.502	0.728
At least one international author	0.373***	0.279	0.000	0.356	0.356	0.864
At least one female author	0.709***	0.628	0.000	0.695	0.695	0.975
At least one top-affiliate author	0.460***	0.315	0.000	0.427	0.427	0.907
Multiple affiliations	0.811***	0.682	0.000	0.797	0.796	0.737
Backward citations Q1	0.090***	0.269	0.000	0.098	0.098	0.935
Backward citations Q2	0.171***	0.249	0.000	0.183	0.182	0.930
Backward citations Q3	0.264***	0.241	0.000	0.280	0.278	0.682
Backward citations Q4	0.475***	0.240	0.000	0.440	0.442	0.626
Life sciences and Medicine	0.439***	0.492	0.000	0.426	0.423	0.564
Mathematics	0.180***	0.189	0.002	0.181	0.182	0.880
Engineering	0.157***	0.199	0.000	0.155	0.155	0.936
Physical sciences	0.506***	0.380	0.000	0.498	0.498	0.915
Year of publication	2011.35***	2010.90	0.000	2011.28	2011.28	0.900

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Appendix U

This appendix reports the covariate balance table pre- and post-matching (Table U1) for the robustness check of section 3.6.6. In this robustness check, we use as a matching procedure the Coarsened Exact Matching (CEM) instead of the *nearest neighbor* Propensity Score Matching.

CEM is a nonparametric procedure that has the advantage to guarantee the balance of the selected set of covariates ex-ante. To do so, the CEM procedure coarsens the support of the joint distribution of the covariates into a circumscribed number of strata. Then, each article is assigned to a unique stratum. We coarsen the joint distributions of the variables listed in Table 12, the same that we use in our main analysis, to create strata. Grant-funded articles and block-funded articles matched together are selected from the same stratum. Strata that do not contain at least one grant-funded article and one block-funded article are discarded. We end up with 6,229 strata. If a stratum contains multiple articles, we impose a 1:1 matching and a grant-funded article is matched with the most similar block-funded article within the same stratum by relying on the *nearest neighbor* Propensity Score Matching (given the same set of covariates of Table 12).

Table U1 shows that the balancing procedure ex-ante allows us to obtain matches where grant-funded articles have all the articles' and authors' characteristics perfectly equivalent to block-funded articles, i.e., the p-values equal 1.000 for all the covariates. The cost is that coarsening the support of the joint distribution of the covariates leads to unmatched grant-funded articles. After the matching, we end up with a sample of 6,229 grant-funded articles (96.7% of the 6,441 grant-funded publications of the main analysis) matched with 6,229 block-funded articles.

Table U1. Means of the articles' and authors' observable characteristics for grant- and block-funded publications, before (Columns 1 and 2) and after (Columns 3 and 4) the Coarsened Exact Matching.

	Sample before CEM			CEM sample		
	(1) **Grant-funded**	(2) **Block-funded**	**p-value**	(3) **Grant-funded**	(4) **Block-funded**	**p-value**
N. of publications	6,441	76,615		6,229	6,229	
Covariates						
Single-author article	0.075***	0.109	0.000	0.074	0.074	1.000
Multi-author article (from 2 to 4 authors)	0.539***	0.483	0.000	0.547	0.547	1.000
Multi-author article (more than 4 authors)	0.386***	0.408	0.001	0.379	0.379	1.000
At least one international author	0.153***	0.279	0.000	0.154	0.154	1.000
At least one female author	0.596***	0.628	0.000	0.599	0.599	1.000
At least one top-affiliate author	0.399***	0.315	0.000	0.398	0.398	1.000
Multiple affiliations	0.693*	0.682	0.058	0.696	0.696	1.000
Backward citations Q1	0.145***	0.269	0.000	0.143	0.143	1.000
Backward citations Q2	0.225***	0.249	0.000	0.226	0.226	1.000
Backward citations Q3	0.278***	0.241	0.000	0.281	0.281	1.000
Backward citations Q4	0.352***	0.240	0.000	0.350	0.350	1.000
Life sciences and Medicine	0.265***	0.492	0.000	0.257	0.257	1.000
Mathematics	0.280***	0.189	0.000	0.270	0.270	1.000
Engineering	0.231***	0.199	0.000	0.222	0.222	1.000
Physical sciences	0.577***	0.380	0.000	0.577	0.577	1.000
Year of publication	2011.31***	2010.90	0.000	2011.30	2011.30	1.000

NOTE: Significance levels at ***p<0.01, **p<0.05, *p<0.1.

Appendix V

This appendix describes how we classify the articles of our sample in their respective fields of research.

To assign an article to a field of research, we refer to the *All Science Journal Classification* (ASJC) scheme that is used by SCOPUS to classify journals under subject areas. Doing so, we classify an article in a research field according to the subject area of the journal where it is published. As stated by SCOPUS, "the classification is based on the aims and scope of the title, and on the content it publishes"[88]. We start from the level of 30 subject areas reported by SCOPUS under the label *Subject Area Classifications*, then we reaggregate them into 5 major fields of research. The way we reaggregate subject areas depends on (i) the affinity among subject areas and (ii) the number of articles in our sample belonging to each subject area.

Table V1 describes how we reaggregate the ASJC subject areas in 5 fields of research. As explained in section 3.3.1 *Data*, we exclude articles classified in Social Sciences from our analysis.

[88] Website: https://service.elsevier.com/app/answers/detail/a_id/14882/supporthub/scopus/~/what-are-the-most-frequent-subject-area-categories-and-classifications-used-in/

Table V1. Reaggregation scheme of the articles' fields of research.

Field of research	***SCOPUS ASJC Subject Area Classifications***
Life sciences and Medicine	Agricultural and Biological Sciences
	Biochemistry, Genetics and Molecular Biology
	Immunology and Microbiology
	Neuroscience
	Pharmacology, Toxicology and Pharmaceutics
	Medicine
	Nursing
	Veterinary
	Dentistry
	Health Professions
Mathematics	Mathematics
	Computer Science
Engineering	Engineering
	Chemical Engineering
	Energy
Physical sciences	Physics and Astronomy
	Earth and Planetary Sciences
	Environmental Science
	Material Science
	Chemistry
Multidisciplinary	Multidisciplinary

GENERAL CONCLUSION

This thesis aims to provide insights to policymakers aspiring to design effective policies for science, and to young and established scholars who want to better understand the determinants of their scientific outcomes. In this thesis, I propose three empirical analyses based on the French case, evaluating what makes a Ph.D. student productive and assessing the effects of different mechanisms to allocate public resources in science.

In this concluding section, I detail the main results and contributions of each chapter, drawing attention to the policy implications.

CHAPTER 1

In the first chapter, I investigate how the characteristics of the environment to which a Ph.D. student is exposed during her doctoral training program relate to the student's scientific productivity. In so doing, I reply to the question: what makes a productive Ph.D. student?

For this study, I create a unique dataset that links information on Ph.D. students to that of their supervisors, peers, university departments, and funding. I cover the entire population of 77,143 Ph.D. students who graduated from French universities in STEM fields between 2000 and 2014. The results show that higher student productivity is associated with having a productive, mid-career, low-experienced, female supervisor who benefits from a national grant. Furthermore, I find that having few productive freshman peers and at least one female peer is positively associated with the student's productivity. I find heterogeneity in the results when breaking down the student population by field of research.

This chapter fills an important gap in the literature: it encompasses in a unique analysis a comprehensive set of relevant training environment characteristics that relate to Ph.D. students' productivity. It considers the characteristics of supervisors, peers, university departments, and funding. Moreover, it covers the entire population of a European country, including 15 cohorts of students in all the STEM fields.

This study offers important insights to Ph.D. students, directors of Ph.D. programs, and policymakers. The hyper-competition in the job market after the Ph.D. training requires students to focus on outcomes that recruiters consider valuable, such as their publication record and scientific network size. With this work, students may learn how to leverage the environmental factors within their specific training to increase their productivity. Directors of Ph.D. programs may also refer to this study to optimize the use of available resources to guarantee effective training programs, increasing the students' future employability. Finally, I provide policymakers with a framework to understand the determinants of effective Ph.D. programs that can be used to design policies that, on the one hand, maximize students' productivity and, on the other hand, allow for an efficient allocation of public resources. For instance, I found that supervisor's grants are associated with increased student productivity, while the fact that the university is awarded government funding for excellence (IDEX) does not relate to student productivity.

Moreover, relying on my econometric estimations, I simulate how French students' productivity varies according to a change in environmental characteristics. The simulation results show that a decrease in the number of peers by one student and a reduction of the average experience of the supervisors by one standard deviation is associated with an increased student's predicted productivity by one publication, one citation, and four additional co-authors.

CHAPTER 2

In the second chapter, I estimate the impact of a competitive funding program for excellence launched by the French government and addressed to universities. Specifically, I assess how the *Initiative D'EXcellence* (IDEX) funding program affects the outcomes of French researchers. Analyzing a panel of 32,947 researchers in STEM disciplines observed between 2006 and 2015, I find that IDEX impacts researchers' outcomes both when the university applies for funding and when the university is awarded funding. Specifically, I find a positive effect of IDEX on the enlargement of the researchers' co-authorship network. Moreover, I show how IDEX indirectly affects the network of researchers in universities that did not apply for IDEX funding but who are connected with awarded researchers. Interestingly, I find an important effect on the internationalization of researchers' networks when universities obtain IDEX funding: both awarded researchers and non-applicant researchers connected to awarded researchers gain around one new additional international co-author per year, relative to researchers in, or connected to, universities non awarded IDEX.

With this chapter, I contribute to the funding literature by analyzing a competitive funding program addressed to universities. Most of the existing studies focus on grants awarded to individual researchers. Moreover, differently from previous studies, I consider a broad set of outcomes of researchers' activities that can be influenced by science funding, such as the quantity, quality, and interdisciplinarity of the publication productivity, interdisciplinarity in the collaborative behavior, scientific collaborations within the laboratory, scientific collaborations within the university, national and international collaborations, patenting, mentoring of Ph.D. students, and fundraising. I also explore two funding effects that emerging literature considers important, namely the effect of applying for funding and the funding spillovers in the co-authorship network.

This chapter offers valuable insights to policymakers considering the recent change of the EU in the rationale for science funding. EU governments are increasingly relying on competitive funding to allocate financial resources to universities at the expense of the more traditional block funding. This new approach emulates that used in the US. However, there is little empirical evidence of its effect on EU universities and it is still unclear whether EU universities would produce similar output to US universities if introduced in a similar context. Many concerns have been raised by scholars: the constraints imposed on universities' agendas, the performance-based competition, and the contractual funding are likely to harm universities' research potential and discourage long-term research projects. Moreover, universities need time to learn and adapt to new mechanisms. Our results show that the effect of the French IDEX competitive funding program is limited to the enlargement of the researchers' network. On the one hand, this finding is encouraging since research collaborations are fundamental in modern science, where high-impact research is increasingly the result of a team effort (Wuthcy et al., 2007). IDEX seems to benefit researchers' international scientific collaborations, one of the main goals of the French government. On the other hand, IDEX does not seem to affect all the other researchers' outcomes. This result is in line with scholars' concerns suggesting how competitive mechanisms for science funding need to be further investigated. A limitation of this study concerns the time window of the analysis. I evaluate the effect of IDEX up to four years after the launch of the funding program. Results are likely to take time to manifest, especially considering the complexity and size of the IDEX program.

CHAPTER 3

The third chapter follows the concerns introduced in the second chapter about the use of competitive incentives in science funding. Specifically, in this chapter, I compare the effects of competitive funding and block funding on the impact of the research produced through the support of these two funding models. I rely on scientific articles' acknowledgments to identify grant- and block-funded articles published between 2009 and 2013. For the grant funding, I focus on the grants distributed by the ANR, the main French funding agency. The ANR aims to promote research excellence in France by granting research proposals through a peer review competitive process. Hence, I implement a probabilistic matching procedure to compare 6,441 grant-funded articles with 6,441 similar block-funded articles.

The main result of this study is that articles resulting from competitive grants receive more citations than articles resulting from block funding in the long run. This finding holds in all disciplines, except in

Mathematics, where articles supported by grant funding are less impactful in the short run and they do not show a different impact in the long run.

The main contributions of this chapter concern the coverage of the entire set of publications resulting from grants awarded by a national funding agency of an EU country, and the methodological approach proposed using a Propensity Score Matching approach to mitigate the selection bias.

For policymakers and scholars, this chapter offers relevant insights. The current debate involving scholars concerns the criticism addressed to the grant funding model appearing to encourage short-term and low-risk research due to the risk-averse behavior of the funding agencies. According to this idea, agencies tend to award research proposals likely to produce results in the short term, instead of funding breakthrough research that needs more time to produce results. Interestingly, in our analysis, we find that grant-funded articles are more impactful than block-funded articles in the long run, but not in the short run. Since grant-funded and breakthrough research share similar citation patterns consisting of higher citations received in the long run (Wang et al., 2017), our results might be explained by the ANR funding agency's effort to support risky research with "high risk/high gain" profile.

BIBLIOGRAPHY

Adams, J.D., Black, G.C., Clemmons, J.R., & Stephan, P.E., 2005. Scientific teams and institutional collaborations: Evidence from US universities, 1981–1999. Research Policy; 34(3), 259–285.

Agarwal, R., Gaule, P., 2020. Invisible Geniuses: Could the Knowledge Frontier Advance Faster? American Economic Review: Insights 2, 409–424. https://doi.org/10.1257/aeri.20190457

Aghion, P., M. Dewatripont, C. Hoxby, A. Mas-Colell, and A. Sapir, 2010. The governance and performance of universities: evidence from europe and the us. Economic policy 25(61), 7–59.

Alberts, B., Kirschner, M.W., Tilghman, S., Varmus, H., 2014. Rescuing US biomedical research from its systemic flaws. PNAS 111, 5773–5777. https://doi.org/10.1073/pnas.1404402111

Alberts, B., 2010. Overbuilding research capacity. Science 329 1257–1257.

Allison, P.D., Long, J.S., Krauze, T.K., 1982. Cumulative advantage and inequality in science. American Sociological Review 47, 615–625.

Allison, P.D., Stewart, J.A., 1974. Productivity differences among scientists: Evidence for accumulative advantage. American sociological review 39, 596–606.

Álvarez-Bornstein, B., & Bordons, M., 2021. Is funding related to higher research impact? Exploring its relationship and the mediating role of collaboration in several disciplines. Journal of Informetrics, Elsevier, vol. 15(1).

Álvarez-Bornstein, B., Morillo, F., & Bordons, M., 2017. Funding acknowledgements in the Web of Science: Completeness and accuracy of collected data. Scientometrics, 112, 1793–1812. http://dx.doi.org/10.1007/s11192-017-2453-4.

Angrist & Jörn-Steffen Pischke, 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Economics Books, Princeton University Press, edition 1, number 8769.

ANR, 2020. Funding research in all its diversity. 2020 annual report. ANR, Information and Communication Division, French National Research Agency. https://anr.fr/fileadmin/documents/2021/ANR-RA2020-en.pdf.

Aristizábal, N., 2021. I bombed the GRE—but I'm thriving as a Ph.D. student [WWW Document]. Science | AAAS. URL https://www.sciencemag.org/careers/2021/06/i-bombed-gre-i-m-thriving-phd-student (accessed 6.17.21).

Arora, A., David, P.A., Gambardella, A., 2000. Reputation and competence in publicly funded science: estimating the effects on research group productivity, in: The Economics and Econometrics of Innovation. Springer, pp. 141–176.

Arora, A., Gambardella, A., 2005. The impact of NSF support for basic research in economics. Ann. Econ. Stat. 91–117.

Autor, D.H., 2003. Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. Journal of Labor Economics, Volume 21, Number 1.

Ayoubi, C., Pezzoni, M., Visentin, F., 2019. The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions? Res. Policy. https://doi.org/10.1016/j.respol.2018.07.021.

Ayoubi, C., Pezzoni, M., Visentin, F., 2017. At the origins of learning: Absorbing knowledge flows from within the team. Journal of Economic Behavior & Organization.

Azoulay, P., Zivin, J.S.G., Wang, J., 2010. Superstar extinction. Quarterly Journal of Economics 42.

Azoulay, P., Ding, W., Stuart, T., 2009. The Impact of Academic Patenting on the Rate, Quality and Direction of (Public) Research Output. The Journal of Industrial Economics. Vol. 57, No. 4, pp. 637-676.

Bandiera, O., Barankay, I., Rasul, I., 2009. Social connections and incentives in the workplace: Evidence from personnel data. Econometrica 77, 1047–1094.

Baruffaldi, S., Visentin, F., Conti, A., 2016. The productivity of science & engineering PhD students hired from supervisors' networks. Research Policy 45, 785–796. https://doi.org/10.1016/j.respol.2015.12.006

Beaudry, C., Allaoui, S., 2012. Impact of public and private research funding on scientific production: The case of nanotechnology. Research Policy, vol. 41, issue 9, 1589-1606.

Belavy, D.L., Owen, P.J., Livingston, P.M., 2020. Do successful PhD outcomes reflect the research environment rather than academic ability? PLOS ONE 15, e0236327. https://doi.org/10.1371/journal.pone.0236327

Benedictus, R., Miedema, F., Ferguson, M.W.J., 2016. Fewer numbers, better science. Nature 538, 453–455. https://doi.org/10.1038/538453a

Bercovitz, J., Feldman, M., 2011. The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. Research Policy 40, 81–93. https://doi.org/10.1016/j.respol.2010.09.008

Biancani, S., L. Dahlander, D. A. McFarland, and S. Smith, 2018. Superstars in the making? The broad effects of interdisciplinary centers. Research Policy 47(3), 543–557.

Bianchini, S., Llerena, P., Öcalan-Özel, S., Özel, E., 2022. Gender diversity of research consortia contributes to funding decisions in a multi-stage grant peer-review process. Humanities and Social Sciences Communications, volume 9, Article number: 195.

Black, G.C., Stephan, P.E., 2010. The economics of university science and the role of foreign graduate students and postdoctoral scholars, in: American Universities in a Global Market. University of Chicago Press, pp. 129–161.

Blundell, R., Dias, M.C., 2009. Alternative Approaches to Evaluation in Empirical Microeconomics. J. Human Resources Summer. vol. 44 no. 3 565-640.

Bol, T., de Vaan, M., van de Rijt, A., 2018. The Matthew effect in science funding. Science, 115 (19) 4887-4890.

Boudard, E., Westerheijden, D.F., 2017. France: Initiatives for Excellence. In: Policy Analysis of Structural Reforms in Higher Education. Palgrave Studies in Global Higher Education. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-42237-4_8

Bozeman, B., Corley, E., 2004. Scientists' collaboration strategies: implications for scientific and technical human capital. Research Policy 33, 599–616. https://doi.org/10.1016/j.respol.2004.01.008

Breschi, S., Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. Journal of Economic Geography, pp. 1–30.

Breschi, S., Lissoni, F., Montobbio, F., 2007. The scientific productivity of academic inventors: new evidence from Italian data. Economics of Innovation and New Technology 16, 101.

Brischoux, F., Angelier, F., 2015. Academia's never-ending selection for productivity. Scientometrics 103, 333–336. https://doi.org/10.1007/s11192-015-1534-5

Broström, A., 2019. Academic breeding grounds: Home department conditions and early career performance of academic researchers. Research Policy 48, 1647–1665. https://doi.org/10.1016/j.respol.2019.03.009

Bruce, A., Lyall, C., Tait, J., Williams, R.A., 2004. Interdisciplinary integration in Europe: The case of the Fifth Framework programme. Futures 36(4):457-470.

Campbell, D., Picard-Aitken, M., Cote, G., Cruso, J., Valentim, R., Edmonds, S., et al., 2010. Bibliometrics as a performance measurement tool for researchevaluation: The case of research funded by the National Cancer Institute of Canada. The American Journal of Evaluation, 31(1), 66–83.http://dx.doi.org/10.1177/1098214009354774.

Carayol, N., Carpentier, E., 2021. The spread of academic invention: a nationwide case study on French data (1995–2012). J Technol Transf. https://doi.org/10.1007/s10961-021-09888-9
Carayol N., Lanoe M., 2018. The Impact of Project-Based Funding in Science: Lessons from the ANR Experience, Post-Print hal-02274256, HAL.
Carayol, N., Matt, M., 2004. Does research organization influence academic production? Research Policy 33, 1081–1102. https://doi.org/10.1016/j.respol.2004.03.004
Carpentier, E., 2022. Two decades of academic invention in France: scope and driving forces. Economics and Finance. Université de Bordeaux, English. NNT : 2022BORD0124.
Cengiz, D., Dube, A., Lindner, A., Zipperer, B., 2019. The Effect of Minimum Wages on Low-Wage Jobs, The Quarterly Journal of Economics, Volume 134, Issue 3, Pages 1405–1454, https://doi.org/10.1093/qje/qjz014
Chenevix-Trench, G., 2006. What makes a good PhD student? Nature 441, 252–252. https://doi.org/10.1038/nj7090-252b
Conti, A., Denas, O., Visentin, F., 2014. Knowledge Specialization in Ph.D. Student Groups. IEEE Transactions on Engineering Management 61, 52–67. https://doi.org/10.1109/TEM.2013.2283039
Cornelissen, T., Dustmann, C., Schönberg, U., 2017. Peer Effects in the Workplace. American Economic Review 107, 425–456. https://doi.org/10.1257/aer.20141300
Correia, S., Guimaraes, P., Zylkin, T., 2020. Fast Poisson estimation with high-dimensional fixed effects. The Stata Journal, 20, Number 1, pp. 95–115.
Corsini, A., Pezzoni, M., Visentin, F., 2022. What makes a productive Ph.D. student? Research Policy. Volume 51, Issue 10, December 2022, 104561.
Cyranoski, D., Gilbert, N., Ledford, H., Nayar, A., Yahia, M., 2011. Education: The PhD factory. Nature 472, 276–279. https://doi.org/10.1038/472276a
Dasgupta, N., Scircle, M.M., Hunsinger, M., 2015. Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. Proceedings of the National Academy of Sciences 201422822. https://doi.org/10.1073/pnas.1422822112
Davies, B., Gush, J., Hendy, S.C, Jaffe, A.B., 2022. Research funding and collaboration. Research Policy. Vol. 51, Issue 2, 104421.
De Boer, H., File, J., Huisman, J., Seeber, M., Vukasovic, M., and Westerheijden, D.., 2017. Policy Analysis of Structural Reforms in Higher Education. Palgrave Macmillan Cham. https://doi.org/10.1007/978-3-319-42237-4.
Defazio, D., Lockett, A., Wright, M., 2009. Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. Res. Policy 38, 293–305. https://doi.org/10.1016/j.respol.2008.11.008
Dehejia, R.H., and Wahba S., 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. The Review of Economics and Statistics (2002) 84 (1): 151–161. https://doi.org/10.1162/003465302317331982.
Delamont, S., Atkinson, P., 2001. Doctoring Uncertainty: Mastering Craft Knowledge. Soc Stud Sci 31, 87–107. https://doi.org/10.1177/030631201031001005
Diamond, A.M., 1984. An economic model of the life-cycle research productivity of scientists. Scientometrics 6, 189–196. https://doi.org/10.1007/BF02016762
Dougherty, K. J., Jones, S. M., Lahr, H., Natow, R. S., Pheatt, L., & Reddy, V. (2016). Looking inside the black box of performance funding for higher education: Policy instruments, organizational obstacles, and intended and unintended impacts. RSF: The Russell Sage Foundation Journal of the Social Sciences, 2(1), 147. https://doi.org/10.7758/rsf.2016.2.1.07
Ebadi, A., Schiffauerova, A., 2015a. How to Receive More Funding for Your Research? Get Connected to the Right People! PLoS One; 10(7), e0133061.

Ebadi, A., Schiffauerova, A., 2015b. On the relation between the small world structure and scientific activities. PloS one; 10(3), e0121129.
Enserink, M., 2008. Will French Science Swallow Zerhouni's Strong Medicine? Science, vol. 322, 28 November 2008, p. 1312.
Eslami, H., Ebadi, A., Schiffauerova, A., 2013. Effect of collaboration network structure on knowledge creation and technological performance: the case of biotechnology in Canada. Scientometrics 97, 99–119.
Etzkowitz, H., 2003. Research groups as 'quasi-firms': the invention of the entrepreneurial university. Research Policy 32, 109–121. https://doi.org/10.1016/S0048-7333(02)00009-4
ERC, 2017. 'European research council, annual report of the erc activities and achievements in 2017'.
Fadda, N., Marinò, L., Pischedda, G. et al. The effect of performance-oriented funding in higher education: evidence from the staff recruitment budget in Italian higher education. High Educ 83, 1003–1019 (2022). https://doi.org/10.1007/s10734-021-00725-4
Falk, A., Ichino, A., 2006. Clean Evidence on Peer Effects. Journal of Labor Economics 24, 39–57. https://doi.org/10.1086/497818
Fernandes, J.D., Sarabipour, S., Smith, C.T., Niemi, N.M., Jadavji, N.M., Kozik, A.J., Holehouse, A.S., Pejaver, V., Symmons, O., Bisson Filho, A.W., Haage, A., 2020. A survey-based analysis of the academic job market. eLife 9, e54097. https://doi.org/10.7554/eLife.54097
Finance Committee of the French National Assembly, 2015. Information report, "la gestion des programmes d'investissements d'avenir relevant de la mission Recherche et enseignement supérieur", http://www2.assemblee-nationale.fr/documents/notice/14/rap-info/i2662/(index)/rapports-information.
Foray, Dominique, 2004. The Economics of Knowledge. MIT press.
Franzoni, C., Stephan, P.E., Veugelers, R., 2022. Funding risky research. Entrepreneurship and Innovation Policy and the Economy 1 (1), 103-133.
Freeman, R., Weinstein, E., Marincola, E., Rosenbaum, J., Solomon, F., 2001. Competition and Careers in Biosciences. Science 294, 2293–2294. https://doi.org/10.1126/science.1067477
Gaulé, P., Piacentini, M., 2018. An advisor like me? Advisor gender and post-graduate careers in science. Research Policy 47, 805–813. https://doi.org/10.1016/j.respol.2018.02.011
Gaulé, P., Piacentini, M., 2013. Chinese graduate students and U.S. Scientific productivity. Review of Economics and Statistics 95, 698–701. https://doi.org/10.1162/REST_a_00283
Gawellek, B., Sunder, M., 2016. The German excellence initiative and efficiency change among universities, 2001-2011. Working Paper, No. 142, Universität Leipzig, Wirtschaftswissenschaftliche Fakultat, Leipzig.
Geuna, A., Rossi, F., 2015. The University and the Economy: Pathways to Growth and Economic Development, Edward Elgar, Cheltenham (2015). ISBN 9781782549482. https://doi.org/10.4337/9781782549499.
Geuna, A., Nesta, L., 2006. University patenting and its effects on academic research: The emerging European evidence. Research Policy. Volume 35, Issue 6, Pages 790-807.
Geuna, A., Martin, B.R., 2003. University Research Evaluation and Funding: An International Comparison. Minerva 41, 277–304.
Geuna, A., 2001. The Changing Rationale for European University Research Funding: Are There Negative Unintended Consequences?, Journal of Economic Issues, 35:3, 607-632.
Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, and M. Trow, 1994. The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies. London: Sage Publications.

Gingras, Y., Larivière, V., Macaluso, B., Robitaille, JP., 2008. The effects of aging on researchers' publication and citation patterns. PLoS One. 3(12):e4048. doi: 10.1371/journal.pone.0004048. Epub 2008 Dec 29. PMID: 19112502; PMCID: PMC2603321.

Gok, A., Rigby, J., & Shapira, P., 2016. The impact of research funding on scientific ouputs: Evidence from six smaller European countries. JASIST, 67(3),715–730. http://dx.doi.org/10.1002/asi.23406.

Golde, F.C.M., 2005. The Role of the Department and Discipline in Doctoral Student Attrition: Lessons from Four Departments 33.

Goldfarb, B., Henrekson, M., 2003. Bottom-up versus top-down policies towards the commercialization of university intellectual property. Research Policy 32, 639.

Grassano, N., Rotolo, D., Moon, J., Lang, F., Hopkins, M.M., 2016. Funding Data from Publication Acknowledgements: Coverage, Uses and Limitations. Journal of the Association for Information Science and Technology, 68(4), 999-1017., SWPS 2016-08, Available at SSRN: https://ssrn.com/abstract=2767348 or http://dx.doi.org/10.2139/ssrn.2767348.

Gulbrandsen, M., Smeby, J.C., 2005. Industry funding and university professors’ research performance. Research Policy 34, 932–950.

Gush, J., Jaffe, A., Larsen, V., and Laws, A., 2018. The effect of public funding on research output: the New Zealand Marsden Fund. New Zealand Economic Papers, 52(2):227–248.

Hackett, E. J. et al., 2021. Do synthesis centers synthesize? A semantic analysis of topical diversity in research. Research policy 50(1), 104069.

Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., 2008. What makes an article influential? Predicting impact in social and personality psychology. Scientometrics, 76(1), 169–185. http://dx.doi.org/10.1007/s11192-007-1892-8.

Heffernan, T., 2021. Academic networks and career trajectory: ‘There’s no career in academia without networks.’ Higher Education Research & Development 40, 981–994. https://doi.org/10.1080/07294360.2020.1799948

Heyard, R., Hottenrott, H., 2021. The value of research funding for knowledge creation and dissemination: A study of SNSF Research Grants. Humanit Soc Sci Commun 8, 217. https://doi.org/10.1057/s41599-021-00891-x.

Horta, H., Cattaneo, M., Meoli, M., 2018. PhD funding as a determinant of PhD and career research performance. Studies in Higher Education 43, 542–570. https://doi.org/10.1080/03075079.2016.1185406

Horta, H., Santos, J.M., 2016. The Impact of Publishing During PhD Studies on Career Research Publication, Visibility, and Collaborations. Res High Educ 57, 28–50. https://doi.org/10.1007/s11162-015-9380-0

Hottenrott, H., Lawson, C., 2017. Fishing for complementarities: research grants and research productivity. Int J Ind Organ 51:1–38.

Hottenrott, H., Thorwarth, S., 2011. Industry funding of university research and scientific productivity. Kyklos 64 (4), 534-555.

Hunt, J., 2010. Why Do Women Leave Science and Engineering? (Working Paper No. 15853), Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w15853

Iacus, S.M., King, G., Porro, G., 2012. Causal Inference Without Balance Checking: Coarsened Exact Matching. Political Analysis, 20, 1, Pp. 1--24. Website Copy at https://tinyurl.com/yydq5enf.

Jacob, B.A., Lefgren, L., 2011. The impact of research grant funding on scientific productivity. J. Public Econ. 95, 1168–1177. https://doi.org/10.1016/j.jpubeco.2011.05.005.

Jaffe, A.B., 2002. Building Programme Evaluation into the Design of Public Research-Support Programmes. Oxf. Rev. Econ. Policy 18, 22–34. https://doi.org/10.1093/oxrep/18.1.22

Johnes, J., 2008. Efficiency and productivity change in the English higher education sector from 1996/97 to 2004/5. The Manchester School 76, 653–674.
Johnes, J., 2006. Data envelopment analysis and its application to the measurement of efficiency in higher education. Economics of Education Review 25, 273–288.
Karaca-Mandic, P., Norton, E.C., Dowd, B., 2012. Interaction Terms in Nonlinear Models. Health Services Research. Methods Corner. Volume47, Issue1pt1, Pages 255-274.
Katz, J.S., Martin, B.R., 1997. What is research collaboration? Research policy 26, 1–18.
Kelchtermans, S., Neicu, D., Veugelers, R., 2022. Off the beaten path: what drives scientists' entry into new fields?, Industrial and Corporate Change, Volume 31, Issue 3, Pages 654–680, https://doi.org/10.1093/icc/dtab054
Lane, J., & Bertuzzi, S., 2011. Measuring the results of science investments. Science, 331, 678–680. http://dx.doi.org/10.1126/science.1201865.
Langfeldt, L., Bloch, C. W., & Sivertsen, G., 2015. Options and limitations in measuring the impact of research grants- evidence from Denmark andNorway. Research Evaluation, 24(3), 256–270. http://dx.doi.org/10.1093/reseval/rvv012.
Lanoë, M., 2018. The evaluation of competitive research funding : an application to French programs. Economics and Finance. Université de Bordeaux, 2018. English. NNT : 2018BORD0363.
Lawson, C., 2013. Academic patenting: the importance of industry support. J. Tech. Transf.38 509–535.
Lee, A., Dennis, C., Campbell, P., 2007. Nature's guide for mentors. Nature 447, 791–797. https://doi.org/10.1038/447791a
Lee, S., Bozeman, B., 2005. The Impact of Research Collaboration on Scientific Productivity. Social Studies of Science 35, 673–702. https://doi.org/10.1177/0306312705052359
Legifrance, 2010. Convention du 23 septembre 2010 entre l'Etat et l'ANR relative au programme d'investissements d'avenir (action « initiatives d'excellence »). https://www.legifrance.gouv.fr/loda/id/JORFTEXT000022851283/
Lempriere, D.M., 2020. What makes a good PhD supervisor? Top tips for managing the student-supervisor relationship. 13.
Lepori, B., Geuna, A., Mira, A., 2019. Scientific output scales with resources. A comparison of US and European universities. PLoS One. 14(10): e0223415. https://doi.org/10.1371/journal.pone.0223415
Levin, S.G., Stephan, P.E., 1991. Research productivity over the life cycle: Evidence for academic scientists. The American Economic Review 81, 114–132.
Lewis, J.M., 2015. Research Policy as Carrots and Sticks: Governance Strategies in Australia, the United Kingdom and New Zealand, Varieties of Governance. Springer, pp. 131–150.
Lewison, G., & Dawson, G., 1998. The effect of funding on the outputs of biomedical research. Scientometrics, 41(1–2), 17–27.http://dx.doi.org/10.1007/BF02457963.
Libaers, D.P., 2012. Time Allocation Decisions of Academic Scientists and Their Impact on Technology Commercialization. IEEE Transactions on Engineering Management 59, 705–716. https://doi.org/10.1109/TEM.2012.2184119
Liénard, J.F., Achakulvisut, T., Acuna, D.E., David, S.V., 2018. Intellectual synthesis in mentorship determines success in academic careers. Nat Commun 9, 4840. https://doi.org/10.1038/s41467-018-07034-y
Liu, W., Tang, L., Hu, G., 2020 Funding information in Web of Science: an updated overview. Scientometrics 122, 1509–1524. https://doi.org/10.1007/s11192-020-03362-3
Long, J.S., McGinnis, R., 1985. The effects of the mentor on the academic career. Scientometrics 7, 255–280. https://doi.org/10.1007/BF02017149
Lovitts, B.E., 2001. Leaving the Ivory Tower: The Causes and Consequences of Departure from Doctoral Study. Rowman & Littlefield Publishers.

Luckhaupt, S.E., Chin, M.H., Mangione, C.M., Phillips, R.S., Bell, D., Leonard, A.C., Tsevat, J., 2005. Mentorship in Academic General Internal Medicine. J Gen Intern Med 20, 1014–1018. https://doi.org/10.1111/j.1525-1497.2005.215.x

Lumley, T., Diehr, P., Emerson, S., & Chen, L., 2002. The importance of the normality assumption in large public health data sets. Annual Review of Public Health, 23, 151–169. https://doi.org/10.1146/annurev.publhealth.23.100901.140546.

Ma, Y., Mukherjee, S., Uzzi, B., 2020. Mentorship and protégé success in STEM fields. Proc Natl Acad Sci USA 117, 14077–14083. https://doi.org/10.1073/pnas.1915516117

Mairesse, J., Pezzoni, M., 2015. Does Gender Affect Scientific Productivity? Revue économique 66 (1), 65–113.

Mangematin, V., 2000. PhD job market: professional trajectories and incentives during the PhD. Research Policy 29, 741–756. https://doi.org/10.1016/S0048-7333(99)00047-5

Mangematin, V., Robin, S., 2003. The two faces of PhD students: management of early careers of French PhDs in life sciences. Sci. and Pub. Pol. 30, 405–414. https://doi.org/10.3152/147154303781780209

Mariethoz, G., Herman, F., Dreiss, A., 2021. The imaginary carrot: no correlation between raising funds and research productivity in geosciences. Scientometrics 126, 2401–2407. https://doi.org/10.1007/s11192-020-03855-1.

Martin, B.R., 2003. The changing social contract for science and the evolution of the university. Science and innovation: Rethinking the rationales for funding and governance. Edward Elgar, Cheltenham, 7-29.

Mas, A., Moretti, E., 2009. Peers at Work. The American Economic Review 99, 112–145.

Mazzucato, M., 2018. Mission-oriented innovation policies: challenges and opportunities. Industrial and Corporate Change, Volume 27, Issue 5, Pages 803–815. https://doi.org/10.1093/icc/dty034.

McHale, J., Harold, J., Mei, J-C., Sasidharan, A., Yadav, A., 2022. Stars as catalysts: an event-study analysis of the impact of star-scientist recruitment on local research performance in a small open economy, Journal of Economic Geography. lbac016, https://doi.org/10.1093/jeg/lbac016

Mejia, C., & Kajikawa, Y., 2018. Using acknowledgement data to characterize funding organizations by the types of research sponsored: the case of robotics research. Scientometrics, Springer, vol. 114(3), pages 883-904.

Merton, R.K., 1968. The Matthew effect in science. Science 159, 56–63.

MESRI, 2019. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000037800979/

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781 [cs].

Mirnezami, S.R., Beaudry, C., Tahmooresnejad, L., 2020. The effect of collaboration with top-funded scholars on scientific production, Science and Public Policy, Volume 47, Issue 2, Pages 219–234, https://doi.org/10.1093/scipol/scz060

Mongeon, P., Paul-Hus, A., 2016. The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics 106, 213–228. https://doi.org/10.1007/s11192-015-1765-5

Moss-Racusin, C.A., Dovidio, J.F., Brescoll, V.L., Graham, M.J., Handelsman, J., 2012. Science faculty's subtle gender biases favor male students. Proceedings of the National Academy of Sciences 109, 16474–16479. https://doi.org/10.1073/pnas.1211286109

Mukherjee, S., Romero D.M., Jones, B., Uzzi, B., 2017. The Nearly Universal Link between the Age of Past Knowledge and Tomorrow's Breakthroughs in Science and Technology: The Hotspot. Science Advances 3(4):e1601315. doi:10.1126/sciadv.1601315.

Nane, G., Larivière, V., Costas, R., 2017. Predicting the age of researchers using bibliometric data. Journal of Informetrics. 11. 713-729. 10.1016/j.joi.2017.05.002.

Nelson, Richard R., 1959. The Simple Economics of Basic Scientific Research. Journal of Political Economy Volume 67, Number 3, Jun., 1959.
Notman, N., Woolston, C., 2020. Fifteen to one: how many applications it can take to land a single academic job offer. Nature 584, 315–315. https://doi.org/10.1038/d41586-020-02224-5
Oancea, A., 2019. Research governance and the future(s) of research assessment. Palgrave Commun 5, 27. https://doi.org/10.1057/s41599-018-0213-6.
OECD, 2019. Education at a Glance 2019: OECD Indicators [WWW Document]. URL https://www.oecd-ilibrary.org/education/education-at-a-glance-2019_f8d7880d-en (accessed 2.4.21).
OECD, 2013. Careers of Doctorate Holders: Analysis of Labour Market and Mobility Indicators (OECD Science, Technology and Industry Working Papers No. 2013/04), OECD Science, Technology and Industry Working Papers. https://doi.org/10.1787/5k43nxgs289w-en
OECD, 2012. Closing the gender gap: Act now. OECD Publishing Paris.
Overington, M.A., 1977. The Scientific Community as Audience: Toward a Rhetorical Analysis of Science. Philosophy & Rhetoric 10, 143–164.
Paglis, L.L., Green, S.G., Bauer, T.N., 2006. Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. Res High Educ 47, 451–476. https://doi.org/10.1007/s11162-005-9003-2
Payne, A, Siow, A., 2003. Does federal research funding increase university research output? Adv Econ Anal Policy 3:1018–1018.
Payne, A., 2002. Do US congressional earmarks increase research output at universities? Sci Public Policy 29:314–330.
Pearson, M., Brew, A., 2002. Research Training and Supervision Development. Studies in Higher Education 27, 135–150. https://doi.org/10.1080/03075070220119986c
Perkmann, M., V. Tartari, M. McKelvey, E. Autio, A. Broström, P. D'este, R. Fini, A. Geuna, R. Grimaldi, A. Hughes, et al., 2013. Academic engagement and commercialisation: A review of the literature on university–industry relations. Research policy 42(2), 423–442.
Pezzoni, M., Mairesse, J., Stephan, P., Lane, J., 2016. Gender and the Publication Output of Graduate Students: A Case Study. PLoS ONE 11, e0145146. https://doi.org/10.1371/journal.pone.0145146
Platow, M.J., 2012. PhD experience and subsequent outcomes: a look at self-perceptions of acquired graduate attributes and supervisor support. Studies in Higher Education 37, 103–118. https://doi.org/10.1080/03075079.2010.501104
Pommier, S., Talby, M., Auffray-Seguette, M., Dalaut, M., Eijsberg, H., Elshawish, P., Muller, H., 2022. Le doctorat en France: Regards croisés sur la formation doctorale (Research Report). Réseau National de Collèges Doctoraux.
Powell, K., 2016. Does it take too long to publish research? Nature News 530, 148. https://doi.org/10.1038/530148a
Rigby, J., 2011. Systematic grant and funding body acknowledgment data for publications: New dimensions and new controversies for research policyand evaluation. Research Evaluation, 20(5), 365–375. http://dx.doi.org/10.3152/095820211X13164389670392.
Rigby, J., 2013. Looking for the impact of peer review: does count of funding acknowledgments really predict research impact? Scientometrics, 94, 57–73.http://dx.doi.org/10.1007/s11192-012-0779-5.
Romano, J.P., Wolf, M., 2016. Efficient computation of adjusted p -values for resampling-based stepdown multiple testing. Statistics & Probability Letters 113, 38–40. https://doi.org/10.1016/j.spl.2016.02.012

Romano, J.P., Wolf, M., 2005a. Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing. Journal of the American Statistical Association 100, 94–108. https://doi.org/10.1198/016214504000000539
Romano, J.P., Wolf, M., 2005b. Stepwise multiple testing as formalized data snooping. Econometrica 73, 1237–1282.
Rong, X., 2014. word2vec Parameter Learning Explained. arXiv:1411.2738 [cs].
Rossello, G., Cowan, R., Mairesse, J., 2020. Ph.D. research output in STEM: the role of gender and race in supervision 57.
Rubin, D.B., 2001. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Health Services & Outcomes Research Methodology 2, 169–188. https://doi.org/10.1023/A:1020363010465.
Rubin, D. B., 1980. Discussion of ‘Randomization analysis of experimental data: the Fisher randomization test’ by D. Basu. J. Amer. Statist. Assoc. 75 591–593.
Rylance, R., 2015. Grant giving: Global funders to focus on interdisciplinarity. Nature News 525(7569), 313.
Sabatier, M., Carrere, M., Mangematin, V., 2006. Profiles of Academic Activities and Careers: Does Gender Matter? An Analysis Based on French Life Scientist CVs. J Technol Transfer 31, 311–324. https://doi.org/10.1007/s10961-006-7203-3
Sacks, F., 2007. Is the NIH Budget Saturated? Why Hasn't More Funding Meant More Publications? The Scientist, November 19.
Sauermann, H., Haeussler, C., 2017. Authorship and contribution disclosures. Science Advances 3, e1700404. https://doi.org/10.1126/sciadv.1700404
Sauermann, H., Roach, M., 2012. Science PhD Career Preferences: Levels, Changes, and Advisor Encouragement. PLOS ONE 7, e36307. https://doi.org/10.1371/journal.pone.0036307
Sarrico, C.S., Rosa, M.J., Teixeira, P.N. et al., 2010. Assessing Quality and Evaluating Performance in Higher Education: Worlds Apart or Complementary Views?. Minerva 48, 35–54. https://doi.org/10.1007/s11024-010-9142-2.
Shibayama, S., 2019. Sustainable development of science and scientists: Academic training in life science labs. Research Policy 48, 676–692. https://doi.org/10.1016/j.respol.2018.10.030
Shibayama, S., Baba, Y., Walsh, J.P., 2015. Organizational design of University laboratories: Task allocation and lab performance in Japanese bioscience laboratories. Research Policy 44, 610–622. https://doi.org/10.1016/j.respol.2014.12.003
Shibayama, S., Kobayashi, Y., 2017. Impact of Ph.D. training: a comprehensive analysis based on a Japanese national doctoral survey. Scientometrics 113, 387–415. https://doi.org/10.1007/s11192-017-2479-7
Shibayama, S., 2011. Distribution of academic research funds: a case of Japanese national research grant. Scientometrics 88, 43–60.
Sinclair, J., Barnacle, R., Cuthbert, D., 2014. How the doctorate contributes to the formation of active researchers: what the research tells us. Studies in Higher Education 39, 1972–1986. https://doi.org/10.1080/03075079.2013.806460
Stephan, P., Veugelers, R., Wang, J., 2017. Blinkered by bibliometrics. Nature 411–412.
Stephan, P., 2012. How economics shapes science. Harvard University Press, Cambridge, Mass.
Stephan, P., Black, G.C., Chang, T., 2007. The small size of the small scale market: The early-stage labor market for highly skilled nanotechnology workers. Research Policy 36, 887–892. https://doi.org/10.1016/j.respol.2007.02.006
Stephan, P., 2006. Wrapping it up in a person: The mobility patterns of new PhDs. Innovation policy and the economy 7, 71–98.

Stephan, P., Levin, S.G., 1997. The Critical Importance of Careers in Collaborative Scientific Research. Revue d'économie industrielle 79, 45–61. https://doi.org/10.3406/rei.1997.1652

Stephan, P., Levin, S.G., 1992. Striking the mother lode in science. the importance of age, place, and time. Oxford.

Tahmooresnejad, L., Beaudry, C., 2019. Collaboration or funding: lessons from a study of nanotechnology patenting in Canada and the United States. J Technol Transf 44, 741–777. https://doi.org/10.1007/s10961-017-9615-7

Tartari, V., Perkmann, M., Salter, A., 2014. In good company: The influence of peers on industry engagement by academic scientists. Research Policy 43, 1189–1203. https://doi.org/10.1016/j.respol.2014.02.003

Taylor, A., Greve, H.R., 2006. Superman or the Fantastic Four? Knowledge Combination and Experience in Innovative Teams. The Academy of Management Journal 49, 723–740.

Tenenbaum, H.R., Crosby, F.J., Gliner, M.D., 2001. Mentoring Relationships in Graduate School. Journal of Vocational Behavior 59, 326–341. https://doi.org/10.1006/jvbe.2001.1804

The Economist, 2021. A growing number of governments hope to clone America's DARPA. The Economist Group Limited. Jun 3, 2021.

Tompkins, K.A., Brecht, K., Tucker, B., Neander, L.L., Swift, J.K., 2016. Who matters most? The contribution of faculty, student-peers, and outside support in predicting graduate student satisfaction. Training and Education in Professional Psychology 10, 102–108. https://doi.org/10.1037/tep0000115

Tonta, Y., Akbulut, M., 2020. Does monetary support increase citation impact of scholarly papers? Scientometrics 125, 1617–1641. https://doi.org/10.1007/s11192-020-03688-y.

Vale, R.D., 2015. Accelerating scientific publication in biology. PNAS 112, 13439–13446. https://doi.org/10.1073/pnas.1511912112

Valero, A., and Van Reenen, J., 2019. The economic impact of universities: Evidence from across the globe. Economics of Education Review, Volume 68, Pages 53-67.

Van Dijk, D., Manor, O., Carey, L.B., 2014. Publication metrics and success on the academic job market. Current Biology 24, R516–R517. https://doi.org/10.1016/j.cub.2014.04.039

Veugelers, R., Wang, J., and Stephan, P.E., 2022. Do Funding Agencies Select and Enable Risky Research: Evidence from ERC Using Novelty as a Proxy of Risk Taking. NBER Working Paper 30320. DOI 10.3386/w30320.

Wahls, WP., 2018. High cost of bias: diminishing marginal returns on NIH grant funding to institutions. Preprint at https://doi.org/10.1101/367847.

Waldinger, F., 2012. Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. The Review of Economic Studies 79, 838–861. https://doi.org/10.1093/restud/rdr029

Waldinger, F., 2010. Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany. Journal of Political Economy 118, 787–831. https://doi.org/10.1086/655976

Wang, J., Lee, Y.N., and Walsh, J.P., 2018. Funding model and creativity in science: Competitive versus block funding and status contingency effects. Research Policy, Elsevier, vol. 47(6), pages 1070-1083.

Wang, J., Veugelers, R., Stephan, P.E., 2017. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. Research Policy. Volume 46, Issue 8, October 2017, Pages 1416-1436.

Wang, J., & Shapira, P., 2015. Is there a relationship between research sponsorship and publication impact? An analysis of funding acknowledgements innanotechnology papers. PloS One, 10(2), Article e0117727 http://dx.doi.org/10.1371/journal.pone.0117727.

Weber K.M., Rohracher H., 2012. Legitimizing research, technology and innovation policies for transformative change: combining insights from innovation systems and multi-level perspective

in a comprehensive ‘failures’ framework. Research Policy, 41, pp. 1037-1047, 10.1016/j.respol.2011.10.015.
Wei, T., Li, M., Wu, C., Yan, X., Fan, Y., Di, Z., Wu, J., 2013. Do scientists trace hot topics?. Sci Rep 3, 2207. https://doi.org/10.1038/srep02207.
Wuchty, S., Jones, B.F., Uzzi, B., 2007. The Increasing Dominance of Teams in Production of Knowledge. Science 316, 1036–1039. https://doi.org/10.1126/science.1136099
Yan, E., Wu, C., & Song, M., 2018. The funding factor: A cross-disciplinary examination of the association between research funding and citation impact. Scientometrics, 115, 369–384. http://dx.doi.org/10.1007/s11192-017-2583-8.
Zhao, D., 2010. Characteristics and impact of grant-funded research: A case study of the library and information science field. Scientometrics, 84,293–306. http://dx.doi.org/10.1007/s11192-010-0191-y.