ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Public research funding and science-based innovation: An analysis of ERC research grants, publications and patents

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Munari, F., Leonardelli, E., Menini, S., Morais Righi, H., Sobrero, M., Tonelli, S., et al. (2024). Public research funding and science-based innovation: An analysis of ERC research grants, publications and patents. RESEARCH EVALUATION, online first, 1-12 [10.1093/reseval/rvae012].

Availability:
This version is available at: https://hdl.handle.net/11585/982895 since: 2025-02-07

Published:

DOI: http://doi.org/10.1093/reseval/rvae012

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

14 September 2026

Public research funding and science-based innovation: an analysis of ERC research grants, publications and patents.

Federico Munari[1], Elisa Leonardelli[2], Stefano Menini[3], Hérica Morais Righi[4], Maurizio Sobrero[5], Sara Tonelli[6] and Laura Toschi[7]*

*Corresponding author: *Via Capo di Lucca, 34, 40126, Bologna, Italy, e-mail:* laura.toschi@unibo.it

Acknowledgements

We gratefully thank the participants of the EPO ARP Workshop 19-20 January 2021 for their valuable comments and suggestions on the paper. We are grateful to Laura Pontiggia, Alexis Michel Mugabushaka and Diego Valerio Chialva of the European Research Council Executive Agency (ERCEA) for their support with data related to patents declared by Principal Investigators of funded projects. Any errors or omissions are our responsibility.

Funding

This work was supported by the European Patent Office Academic Research Programme [2018/8401].

[1] Professor, Department of Management, University of Bologna, *Via Capo di Lucca, 34, 40126,* Bologna, Italy
[2] Researcher, Fondazione Bruno Kessler, Via Sommarive, 18, 38123, Trento, Italy
[3] Researcher, Fondazione Bruno Kessler, Via Sommarive, 18, 38123, Trento, Italy
[4] Researcher Fellow, Department of Management, University of Bologna, *Via Capo di Lucca, 34, 40126,* Bologna, Italy
[5] Professor, Department of Management, University of Bologna, *Via Capo di Lucca, 34, 40126,* Bologna, Italy
[6] Head of Unit, Fondazione Bruno Kessler, Via Sommarive, 18, 38123, Trento, Italy
[7] Associate Professor, Department of Management, University of Bologna, *Via Capo di Lucca, 34, 40126,* Bologna, Italy

Public research funding and science-based innovation: an analysis of ERC research grants, publications and patents.

This study aims to advance the discussion around the linkages between public funding for scientific research and new technology development. We develop and test a methodology that matches patents to scientific publications, the latter of which stemming from publicly funded research. We specifically focus on projects that were funded by the European Research Council (ERC) in the Life Sciences and Physical Science and Engineering sectors during the FP7 Programme. We also compare this method's results with the patents directly reported by the PIs of ERC-funded projects at the end of their grants, finding that the two methods are complementary. In addition to documenting important technological diffusion effects generated by ERC-funded research, our results highlight some factors that shape the relationships between patents, publications, and grants, such as project type, age, duration, scientific domain and number and quality of associated publications. Therefore, our findings present significant policy implications for funding agencies, universities, TTOs, and policymakers who want to monitor the technological outcomes of public research investments.

Keywords: public research funding; patents; scientific publications; technology transfer

JEL: O32; O34; O38

Introduction

Innovation is a crucial tool in searching for improved conditions for industry and society by fostering new products and services, production processes, and business models. Consequently, the literature on the economics of science and innovation has invested great effort in understanding the impact of public funding for research activities that strive to expand the scientific and technological frontier (Auranen and Niemine, 2010; Beaudry and Allaoui, 2012). Generally, a variety of indicators based on scientific publications and the citations they receive are used as proxy to measure the impact of public research funding in the advancement of science. However, there is increasing recognition that we need to also analyze the outcomes in terms of new technologies, commercial applications, and societal results. In this vein, a

foundational stream of literature has spent years studying the links between scientific articles and patents—the latter of which are considered an intermediate output of research commercialization (Jaffe, 1989; Jaffe and Trajtenberg, 1998).

In this context, a promising line of research has recently emerged that traces patent citations to publicly-supported scientific publications (Azoulay et al., 2019; Fleming et al., 2019; Li et al., 2017). These studies leverage advancements in text analytical techniques alongside the increasing availability of large-scale patents, publications, and grants databases. In short, these authors have revealed the linkages between public research funding and technology development by analysing patents' scientific citations.

Although this approach is promising and opens up a wealth of new applications, it still presents several open questions. First, previous studies have focused on specific grant schemes, often within a particular scientific field—mostly the Life Sciences. Consequently, there is a limited understanding of the influence exerted by the types of competitive grant schemes and by the different scientific fields on the reliance of patents on science. To address this gap, we compare different funding schemes introduced by the ERC and analyze different scientific fields (both life science and physical and engineering) and capture differences in their ability to support science able to inspire the generation of new patented technologies. Second, many existing studies revolve around US patents and institutions while neglecting other institutional environments. In our work, we address this gap by replicating the approach in the European context, especially considering the significant differences not only between the US and Europe, but also among European countries (Dosi et al., 2006). Third, the literature mainly focuses on patents' outcomes, as identified through extensive, large-scale analyses based on patent-publication-grant matches. We enlarge this perspective by analysing also those patents directly reported by the coordinators of such grants and comparing the two channels (self-reported patents and citing patents). Such an analysis may help identify complementarities or overlaps

between the different channels and techniques used to identify influenced patents—and thereby fostering a more rigorous and comprehensive assessment of the reliance of patenting activity on publicly-funded science.

Addressing these gaps, our research introduces a novel method of analyzing the relationship between scientific and technological productivity (Pisano, 2010; Perkman and Walsh, 2009; Rake, 2021). We intend to comprehensively detail how patented inventions inspired by publicly-funded science follow development processes characterized by a long-term time span. Indeed, these innovations generally emerge several years after receiving the funding and flow through different channels: from patents directly developed by the research project to scientific publications that inspire later patents.

To this end, we developed and tested a new methodology for identifying granted patents that rely on publicly funded scientific publications. Our analyses are based on projects funded by the European Research Council (ERC) during the FP7 Programme, specifically in the Life Sciences (L.S.) and Physical Science and Engineering (P.E.) sectors where the propensity to patent is higher (with respect to Social Science and Humanities). We identified the granted patents that cited scientific publications stemming from ERC-funded research projectsand compared them with those directly reported by the Principal Investigators (PIs) at the end of their grants. Our approach involved four steps: We first identified all ERC-funded FP7 projects in these sectors. We then used the E.U. CORDIS and Scopus platforms to uncover the scientific publications stemming from these projects. In the third step, we traced the set of granted patents that cited those publications, using a matching algorithm that pulls data included in the NPL (Non-Patent Literature) section of Patstat. Finally, we compared these results with the set of patents reported to ERC by the PIs at the end of the funding period.

The focus on projects funded by the ERC offers several advantages. First, given the articulation of ERC grants, we can analyze the impact of different types of grant schemes,

differentiating between frontier research vs. translational research grants (e.g., Proof-of-Concept Grants) as well as grants for early-stage vs. advanced-stage researchers. Moreover, we can examine differences in patent outcomes by scientific sectors (specifically between L.S. and P.E.). Second, we can combine and compare evidence from several European countries against previous results that were solely confined to the United States. Third, by using ERC's data on PIs' declared patents as a comparison, we can address the short-term impact of the knowledge developed by these research projects (as captured by the patents declared to the ERC at the end of the project) along with their long-term potential to inspire technology generation (as captured by subsequent patents citing the project publications).

Following this track, we identified 3,697 FP7 ERC grants awarded in the L.S. and P.E. domains, which self-reported 951 patents and generated 86,630 scientific publications. Through our matching method, we identified 12,918 unique patents citing those publications. Overall, about 30% of ERC-funded projects published papers that were subsequently mentioned by granted patents. This share is significantly higher than that of projects declaring a granted patent to the ERC, around 4%. In addition, there is relatively limited overlap among the results provided by the two approaches; in other words, few patent-linked projects were simultaneously present on both sets. Thus, our analysis shows that the two approaches can be considered complementary, as they capture patenting outcomes in different ranges of time (medium- and long-term versus short-term).

Finally, we delineate some of the factors that are predictive of a grant-patent linkage. Our matching method shows that projects characterized by higher age (assessed as the difference beyween the year 2020 and the starting year of the project), duration, budget level, and number and quality of associated publications have a higher likelihood to be cited by a patent. On the other hand, the matching method is less effective in capturing patents linked to projects with a translational nature, given that they tend to be more focused on pursuing

valorisation activities, rather than on generating new scientific publications. Therefore, our findings underscore the benefits and limitations of a large-scale approach to monitoring public research investments' technological results. From this, we derive significant policy implications for funding agencies, universities, TTOs, and policymakers.

This paper has five sections. The first one presents a literature review on the importance of publicly funded research for technological productivity and the recent attempts to measure its impact by linking research grants and patents through publications. Section two describes the ERC context, while section three details the methodology and databases. The fourth section discusses the results and section five offers our conclusions.

1 Literature Review

1.1 The importance of public investment on technology development

There is a long tradition of analyzing the direct and indirect effects of public investments in research (Griliches, 1992; Jaffe, 1989; Munari and Toschi, 2021). In terms of research on technology development, the literature has used different approaches to measure its impact. The first one traces the influence of public research investment on technology generation using companies' information about their capacity to generate new technologies (Mansfield, 1991; Cohen et al., 2002). Another approach uses patent information (Jaffe, 1989; Jaffe et al., 1993). Applying a knowledge production function, Jaffe (1989) investigates the existence of spillovers from university research to commercial innovation, exploiting panel data at the state-level on university research funding and corporate patents. In a similar vein, Jaffe et al. (1993) used patents' citations of scientific publications to assess how public research funding leads to knowledge spillovers across geographic regions.

An emerging stream of this literature has focused on research grants as the unit of analysis and their resulting publications as the main outcome measure (Wang and Shapira,

2011; Beaudry and Allaoui, 2012). This has opened a promising line of research whereby patents serve as a proxy of these projects' technological impact (Azoulay et al., 2015; Grassano et al., 2017), as we discuss in the next section.

1.2 The linkage between research grants and patents

With the growing availability of data on specific grants, a line of research has emerged around the impact of public funding on patent generation. This stream seeks to reconstruct the links between these research projects, their publications (e.g., through acknowledgments), and patents citing those publications.

The first step is to investigate the link between patents and publications using the Non-Patent Literature (NPL) information available on those patents. This approach uses the patents' references to publications that inspired the creation of new technological inventions. These citations can serve as a proxy for the knowledge flow between science and technology (Jaffe et al., 1993; Tijssen et al., 2000; Callaert et al., 2014; Azoulay et al., 2015; Veugelers and Wang, 2019). On this regard, an interesting contribution is offered by Callaert et al. (2014), who analysed 33 Belgian inventors concerning the importance of non-patent references included in the front-page of patent documents. The interviews revealed that scientific literature inspired 67% of the patents with front-page scientific references (versus 33% of the patents with no scientific citations). This result confirms that the existence of scientific citations on patents increases the possibility that scientific results inspired the development of a new technology.

The second, and more recent, step of this approach links specific research grants to new patented technologies via cited publications. A seminal paper in this area is the work of Azoulay et al. (2015), who developed a method for linking NIH research funding to USPTO patents owned by pharmaceutical and biotechnology firms. In their novel methodology, the authors used data from NIH grants to identify scientific publications by their acknowledgment data and

then linked those publications to patents through USPTO bibliometric references. This study showed that 43% of the projects generated at least one publication linked to a patent. They also showed that the increment in public research investment is directly related to increased private R&D investment and valuable technologies. These results are also confirmed in other settings by Du et al. (2019), Jefferson et al. (2018) and Poege et al. (2020). Similarly, De Rassenfosse et al. (2019) focused on the innovation outcomes stemming from procurement contracts in order to clarify the impact of government demand (rather than scientific research projects) on technological development. The authors found that 1.5% of the procurement contracts originated at least one patent; however, these contracts corresponded to 36% of the overall contract value.

Li et al. (2017) introduced the distinction between direct and indirect patents, defined respectively as those citing a specific grant in their acknowledgment section and those with a citation to a publication generated by those grants. They found that about 10% of NIH grants directly generated a patent, and 30% generated articles that were subsequently cited by patents.

Fleming et al. (2019) showed that government-supported science papers increased their impact on private companies' technology development rather than government-supported patents. In particular, their study demonstrated that the average percentage of US federal research patent citations by private companies' patents decreased from 2003 to 2017. On the other hand, the publications' citations increased within the period investigated (from 2003 to 2017). This result underscores the increasing importance of public science in inspiring innovation and private technology development.

Tables 1 and 2 in the Online Supplementary Materials section report a summary of these studies, including the data sources, the methodological approach for linking patents to data, and the main results.

Although the literature is increasingly interested in the practice of connecting research grants, scientific articles, and patents, there are a few areas of improvement that could make the results more accurate. For instance, Grassano et al. (2017) showed a disparity between the number of papers funded by public grants and the number of publications with a declaration of funding support as reported in databases like Web of Science (WoS) and PubMed. In the same vein, Callaert et al. (2014) argued that an important share of the papers cited on patent documents are often used as background information rather than being a central component of new inventions. In addition, some authors have claimed that this approach of identifying patented inventions via citations to publications can underestimate the scientific contribution of said projects to technological development, given that they do not capture all channels of knowledge that flows between science and technology (Roach and Cohen, 2013). Even though, the analysis of publications cited by patents provides a proxy of the linkages existing between public research findings and the development of new technologies (Roach and Cohen 2013; Azoulay et al., 2015).

Considering these limitations, we thus sought to develop a rigorous and accurate methodology that could assess the presence of linkages between research grants and the generation of new patented technologies. In contrast to existing research, which has mainly revolved around specific US funding agencies specialized in the life sciences, we focus on the patent-related outcomes of ERC research grants in Europe comparing two macro-sectors (i.e. Life Sciences, and Physical Science and Engineering). In this way, we test a new methodology for linking research grants, scientific publications, and patents, and we compare the results with those obtained by patents directly declared by the PIs. Finally, we analyze a set of relevant determinants of the likelihood to observe a patent-grant match.

2 The Setting: Research Grants from the European Research Council

Our paper focuses on data related to the ERC FP7 Programme. The European Research Council (ERC) was established in 2007 to support the highest-quality frontier research across all fields and thereby strengthen the European research system. The ERC provides individual long-term grants for ground-breaking, high-risk/high-gain research through competitive funding. A particular feature of ERC funding grants is that the projects are assigned to the PI and not to the Host Institution, which confers the researchers with greater mobility (i.e., they can continue the research in the event of a transfer).

The Council offers five main types of grants that represent the core of its activities: Starting Grants (STG), Advanced Grants (ADG), Consolidator Grants (COG), Proof of Concept Programme (PoC), and Synergy Grants (SYG). Through this division, the ERC can support early-career researchers while also galvanizing established researchers to develop their ideas and contribute to the wider scientific field. The Council aggregates various disciplines into three macro fields: Life Sciences, Physical Science and Engineering, and Social Sciences and Humanities. Table 2 in the Online Supplementary Materials section describes the main characteristics of these grant categories.

Beneficiary projects are selected based on a high-quality, international peer-review process—with the sole criterion being scientific excellence. Since 2007, ERC has funded more than 14,000 research projects[8] across all of its grant categories. Our study considers only ERC projects funded under the 7th Framework Programme, which existed from 2007 to 2013[9]. Over that period, the ERC financed over 4,500 FP7 projects with a budget of €7.5 billion.

[8] This information is available at the ERC website (https://erc.europa.eu/project-statistics/project-database).

[9] We did not include ERC projects funded under the more recent H2020 framework programme (2014-2020) in order to ensure there was sufficient time to observe the patenting effects after the projects' conclusion. Considering the duration of ERC projects (up to 5 years), a large share of H2020 ERC projects were still ongoing, or in their very early phases, at the time of our study.

3 Methods and data-sources

3.1 Construction of the dataset

Figure 1 illustrates the four steps of our method for search patents that cited scientific publications, which were themselves the result of ERC-funded research projects.

--- Figure 1 about here ---

3.1.1 First step: identifying FP7 ERC Projects

As a first step, we gathered the list of projects awarded with an ERC grant. We initially collected all projects funded by the ERC in all Programmes (FP7 and H2020) and sectors (Life Sciences, Physical Science and Engineering, Social Sciences and Humanities). We filled the list by collecting and aggregating data from different sources: ERC website, CORDIS and European Union Open Data Portal[10]. This step identified 4,555 ERC FP7 projects.

For our final sample, we excluded Programme grants awarded in the Social Sciences and Humanities domain, given the extremely low propensity to patent in this field. We also excluded projects funded by the Synergy Grant, as they are larger (about € 10 million instead of € 1.5-2.5 million) and longer projects (6 years instead of 1.5-5 years), usually developed by multiple fields and research teams (up to four PIs instead of one). These characteristics make them different in nature if compared to the other projects included in our sample and they represent a very small proportion of the total number of funded projects (0.53%). Thus, our final sample included 3,697 FP7 projects from the Life Sciences (L.S.) and Physical Science and Engineering (P.E.) field. See Online Appendix A for additional information about this step.

[10] European Union Open Data is a web portal that provides access to public data published by European Union institutions; available at https://data.europa.eu/en.

3.1.2 Second step: linking ERC projects to scientific publications

We used two different and independent sources to capture relevant information about those ERC projects: CORDIS and Scopus[11]. Collecting data from different platforms allowed us to capture a larger number of publications, despite a partial overlap between the two sources[12]. We matched the publications from each source based on the title and/or the Digital Object Identifier (DOI). After merging and cleaning the data[13], we arrived at 86,630 publications linked to ERC FP7 projects from the L.S. and P.E. domains. Online Appendix B elaborates on the process of collecting and aligning publications between CORDIS and Scopus.

3.1.3 Third step: linking scientific publications to patents

After defining the dataset, we used the Patstat database to trace the patents that cited publications generated by the grants[14]. First, we matched those publications with the Non-Patent Literature (NPL) field from Patstat, i.e., the dedicated field that identifies scientific publications that are cited on the patent's front page or during the examination process[15]. We relied on three criteria to match the NPL to the corresponding scientific publication: the same first author's surname, the same title, and the same DOI (see Online Appendix C for more details on our matching methodology). We selected those with at least two matches, and manually checked the publications that only featured the same title or DOI.

---

[11] Elsevier's database of publications.
[12] In our final dataset, we only used publications existent on Scopus' database published after the beginning of the project.
[13] The methodology used on this step is detailed in Online Appendix B at the end of the paper.
[14] Patstat is a database released by the European Patent Office (EPO) that provides access to patent data from more than 40 patent authorities worldwide, including the World Intellectual Property Office (WIPO).
[15] Patstat presents metadata for NPL papers in two configurations. There are structured entries, with a separate field for every piece of information (e.g., title, authors, DOI, or year); those comprise about nine million entries. Conversely, unstructured entries—where all information about the publication occupies a single free-text field—constitute the majority (about 35 million).

After linking the NPL table entries to ERC grants, we further combed the Patstat database to characterize the citing patents. For each publication found in the previous match, we gathered additional data regarding the citing patents, such as application number, grant process status, year of application, patent family information, and other additional information. Moreover, we distinguished so-called "direct" patents from "indirect" patents: the former are patents citing previous ERC-funded publications with the same project PI (among the applicants and/or inventors) and the Host Institution (Coordinator) or a partner institution among the applicants. The assumption here is that such patents are a direct output of the project. Meanwhile, "indirect" patents are those assigned to different applicants and inventors; we consider these to be a proxy of the broader diffusion in the technological domain of the scientific knowledge generated through the project.

Through this method, we were able to link each patent to a respective project via the latter's publication(s). Our dataset contains 15,483 links between research projects and patents, considering 12,918 unique patent documents citing FP7 projects' publications, 6,670 of which were granted patents at the time of the study. These patents stem from different patent offices: USPTO[16] patents were the most predominant (83.3%), followed by EPO[17] (7.4%) and CNIPA[18] (5.6%)[19]. However, a deeper analysis considering the patent families identified, we found out that 47% of the granted patents had at least one European Patent on the family.

We also differentiated direct and indirect citing granted patents. As explained before, direct citing patents are those patents where the PI is one of the inventors and/or applicants and the Host Institution or Partner Institution is one of the applicants. On the other hand, indirect patents

[16] United States Patent and Trademark Office
[17] European Patent Office
[18] China National Intellectual Property Administration.
[19] The concentration of USPTO Patents in comparison to other patent offices is due to the more rigid North American legislation regarding the state of art declaration requirement (Michel and Bettles, 2001).

are those developed by other inventors and institutions. From the 15,483 links between research project and patents, 887 are registered as direct citing patents distributed among 313 FP7 projects.

3.1.4 Fourth step: a comparison with the data declared to the ERC by the PIs

In the fourth step, we compared the results obtained from the matching method with the set of patents reported by the projects' PIs in the intermediate and final technical reports submitted to the ERC. Our goal here is to better understand how these two approaches overlap or complement each other. To this end, we used information directly provided by the ERC regarding the patents reported by each FP7 project. The ERC's internal database had 1,316 patent records: 1,030 declared from FP7 projects and 286 from more recent (and in many cases, still on-going) H2020 projects. From that list, we used the patent application number to match them to Patstat database in order to gather information about their application year, legal status, inventors, applicants, patent family, technology domain[20], and NPL citations. From this search, we traced 1,157 unique patents distributed across 508 projects, including both FP7 and H2020 projects (one patent was linked to more than one project). Restricting our focus to FP7 projects from the L.S. and P.E. sectors, we arrived at 951 unique patents distributed across 351 projects. From these patents, we identified 352 granted patents. In terms of regional origin, 50% of the patents came from the US Patent Office, 27.6% came from the European Patent Office, and 2.6% came from the Japanese Patent Office.

3.2 Empirical Strategy

[20] For each patent, we gathered information about the International Patent Classification (IPC) and Cooperative Patent Classification (CPC).

To better understand how complementary the two methodologies (i.e., the matching vs. reporting method) are, we used each method to uncover the factors that shape a project's probability of being cited by at least one patent.

To that end, we employed a Probit specification whose binary outcome represents projects with (or without) citing patents. We then conducted two sets of analysis that represent the two methods and compared their results. In both models, we considered two main groups of explanatory factors: (i) characteristics at the project level and (ii) the scientific production of the project team. In terms of project characteristics, we looked at the type of project (Advanced, Consolidator, Starting or Proof-of-Concept grants), the scientific sector (L.S. vs P.E.), the project age and duration, the received funding amount, and the number of institutions involved in the project. We also looked at the number of scientific publications generated by the project and the number of scientific citations they received. Finally, we controlled for geographical effects through a set of dummies that captured the location of the project's European host institution. Table 1 provides a more detailed description of the variables included in our regression analyses.

--- Table 1 about here ----

4 Analyses and Results

4.1 Distribution of patent outcomes by project characteristics

Table 2 shows the summary statistics of the variables included in our econometric specifications, while Table 3 presents the correlations between those variables for the 3,697 projects in our sample. Table 4 compares projects in the L.S. and P.E. domains in terms of (i) the number (and percentage) of projects whose publications were cited by at least one patent and (ii) the number (and percentage) of projects declaring at least one patent in the final report for the ERC by sector and grant.

--- Tables 2, 3 and 4 about here ----

Overall, 29.11% of the projects that produced publications were cited by subsequent patents. By comparison, this value was only 4% when relying on PIs' declaration to the ERC that the project had generated a patent. This discrepancy offers preliminary evidence for the amount of time needed to translate research outcomes into patentable activities. Thus, there are significant technological spillover effects beyond what can be captured from merely analyzing patents declared by the PIs of ERC-funded projects. Usefully, our matching method revealed patent citations indexes that are similar to those found in previous research on NIH funding grants (Li et al., 2019). In other words, this method captures patents that are inspired by the knowledge generated several years after the conclusion of a project, thus serving to map influence in the medium- and long-term. On the contrary, the method based on reported patents is limited by a short-term focus (PIs must typically submit their final technical reports about six months after the project's completion) and potential underreporting by the involved researchers—the latter issue having already been documented by research based on US biomedical data (Rai and Sampat, 2018). Furthermore, we observed a limited overlap between the two groups: Only 95 (8.83%) of the projects with a linked patent (citing their publications) also reported a patent. Conversely, only a small fraction of patents reported at the ERC include a citation to previous publications stemming from the same project. We interpret this result as research teams deciding to patent before publishing in order to satisfy the novelty requirement for patenting[21]. From a methodological point of view, this important finding suggests that focusing the

[21] In order to validate this interpretation, we performed three interviews with PIs of ERC-funded projects and with a representative of a Knowledge Transfer Office responsible for patenting activities. All the interviews confirmed that there is a close collaboration between KTO and PIs, and that the latter receive the suggestion to patent first and then publish if there is a possibility of realizing the patent.

investigation exclusively on the method based on patent-publications matches risks underestimating the contribution related to patents generated as a direct output of research projects, for instance those developed in the course of the project and filed in the short term by the PI and her research team.

Looking more closely at the data, we observed similar probabilities between the two focal macro-domains in terms of scientific publications being cited by patents (29.29% for L.S. versus 28.96% for P.E.). This scenario changes slightly when we analyze the results based on projects with reported patents, with P.E. projects leading to a slightly higher share of reported patents (4.50% versus 3.37% for L.S.). This result could be partially explained by the Life Sciences field needing more time to complete an innovation development process.

Table 4 also shows the distribution of patent outcomes in terms of ERC grant type—for both the total sample and for our two focal sectors separately. We observed the highest likelihood of patents citing project's publications among Advanced Grants (33.86%), followed by Starting Grants (30.60%) and Consolidator Grants (9.45%). Interestingly, Advanced Grants (those assigned to researchers with higher seniority and an excellent scientific track-record) were especially likely to have a publication cited by patents in the L.S. domain, but we did not observe any major differences in the P.E. domain. Note that Consolidator Grants were introduced later than the other two types (2012 compared to 2007); thus, their lower percentage can be attributed, at least in part, to having less time to generate publications.

We observed a similar distribution among projects reporting a granted patent in the final reports, albeit at the lower percentage levels illustrated before. We did not include Proof-of-Concept (PoC) projects information for two reasons. First, such projects followed different internal reporting rules in the FP7 Programme, also in respect to reporting information on patents generated. . Second, based on the matching method, PoC programmes produced a small share of projects linked to patents citing their direct publications (only 6.37%). This result may

seem surprising given the translational nature of PoC projects, which generally build on previous ERC Frontier Research Grants and have a more applied nature compared to the other types of grants. However, POC projects have few direct publications per project, which can be explained by the facts that (i) the majority of the publications are linked to the previous corresponding initial Frontier Research Grant (i.e. Starting/Advanced/Consolidator) on which they build upon and (ii) the shorter duration of PoCs (18 months) in respect to the other project types (5 years). Therefore, given that translational research projects, such as PoCs, develop knowledge previously generated through other grants (starting, advanced, …), both types of research are not easily comparable looking at patent-to-publications matches. Thus, another contribution of our analyses is the awareness that the patent-to-publications matching method is likely to significantly underestimate the impact of PoC projects in terms of subsequent patents.

In order to better characterize our sample of projects, Online Appendix D reports the results of additional analyses that compared projects that generated publication cited by patents with projects that did not receive any patent citation. In order to compare the impact of science and technology produced by these research projects, Online Appendix E also compared projects that reported at least one patent and those that did not report a patent in the final report. The results show that projects with higher patent citations are older and longer. These results are similar to those founded by the literature (Rai and Sampat, 2012; Azoulay *et al.,* 2019) and demonstrates that novel science can be risky and most of times need further development in order to increase their potentiality to impact science and technology development (Veugelers and Wang, 2019).

Other result pointed out in Appendices D and E is that projects cited by patents and/or generated a new patent present a larger budget and higher number of publications. In line with

this result, Azoulay *et al.* (2015) showed that a $10 million increase in funding result on 2.3 additional patents.

4.2 Regression Analyses

According to the correlation matrix presented in Table 3, the variables in our econometric models were not highly correlated and thus should not cause multi-collinearity problems. We further supported this result by analysing the mean variance inflation factors, which were below the value of 3.04.

Table 5 shows the results of our regressions. Models 1 and 2—which predict the probability that a project will generate publications that will later be cited by granted patents—cover all FP7 projects in the L.S. and P.E. domains (3,697 observations). Models 3 and 4, instead, estimate the same econometric specifications, but focus on the probability that a project will generate at least one granted patent, i.e. patents reported by the PI in the final report[22]. For each dependent variable, we performed two different analyses: Models 1 and 3 assessed the impact of specific project characteristics (i.e., type of Programme, age, number of partners, budget and duration), while Models 2 and 4 also included the effect of the project's scientific productivity in terms of quantity (i.e., number of publications) and quality (i.e., number of citations of said project's publications).

--- Table 5 about here ----

Notably, several of our variables seemed to influence the probability that a project had at least one publication cited by a granted patent. Consider Model 2: Projects funded by a Starting Grant were relatively more likely (than Advanced and Consolidator grants) to inspire

[22] Models 3 and 4 do not include the projects funded through PoC grants because this category has different reporting rules in the FP7 Programme and thus lacks this type of information, as discussed in the previous section. As a result, such Models reflect a total of 3,540 observations.

subsequent patents; this difference was statistically significant at the 1% level (Starting Grants were used as a baseline case of the categorical variable capturing the type of funding). Moreover, projects in the Life Sciences were more likely than projects in Physical Engineering to generate publications that were then cited by subsequent patents (0.167, $p<0.01$). We observed a similar pattern regarding project age: older projects were more likely to register citing patents than younger ones (0.206, $p<0.01$) because they had more time to develop. The same was true for projects with a higher budget (0.639, $p<0.01$), suggesting that the development of technological outputs is a costly activity. Meanwhile, an increase in the duration of a project translated into a lower probability of being cited by a patent (-0.188, $p<0.01$). Of course, it is still possible that these projects can be cited by future patents. Moreover, the number of institutional partners involved in a project has a negative and significant effect on the investigated dependent variable (-0.158, $p<0.05$). This result may be due to conflicting IP arrangements between collaborating institutions or the difficulty of attributing credit to the involved parties[23]. Finally, regarding the scientific productivity of the project team, we found that projects with a higher number of publications and a higher number of citations have an increased probability of being cited by at least one granted patent ($p<0.01$). In other words, projects with higher-quality research outcomes are more likely to inspire subsequent patents.

Models 3 and 4—which focus on patents reported by the PI to the ERC at the end of the funding period—showed results that largely parallel, but also diverge from, the previous analyses. In particular, the projects funded through the Starting Grant Programme were more likely to report patents as a final outcome than Advanced Grant projects (0.264, $p<0.01$) (there were no statistically significant differences with respect to Consolidator Grants). On the other

[23] We thank one of the Reviewers for suggesting this interpretation of the results.

hand, projects in the L.S. domain were less likely to report patents than those in the P.E. domain—a result that runs counter to the analyses shown in Model 2. This finding may be due to the patenting process being longer and more complex for L.S. than P.E. As observed in the previous analysis, project age and project budget were positively related and statistically significant at the 1% level. Moreover, the results regarding project duration and the scientific productivity of the team were not statistically significant.

Finally, we performed additional robustness checks in order to further validate our results. In the regressions discussed above, we considered as dependent variable a dummy variable to investigate the probability of a project to be cited by a granted patent. However, as some projects in our sample have multiple patents, we also used a negative binomial model in order to test if our results hold when considering the probability of obtaining a higher number of direct/indirect patents. The results did not present significant differences. These additional analyses are available upon request.

5 Conclusions

This study sought to discuss to which extent public funding for scientific research inspires the development of new patented technologies. Following the technical advancements in this rapidly growing field of literature, we developed an innovative methodology for linking research grants to patents through publication citations. We focused on research grants that were awarded by the European Research Council within the FP7 Programme, and specifically in the Life Sciences and Physical Science and Engineering domains (over a period ranging from 2007 to 2013). We chose the FP7 Programme in order to ensure that the research projects would have sufficient time to produce measurable impact (De Rassenfosse et al., 2019; Veugelers and Wang, 2019).

Our results confirm that ERC-funded research present important and broad connections with subsequent patented technologies: the share of research projects in our sample whose publications were cited in at least one granted patent (29%) was significantly higher than the share of projects that reported the development of a granted patent to the ERC (4%). These results align with those reported for the National Institute of Health in the US (Li et al., 2019). Thus, there appears to be a gap between what research teams formally report at the end of a project and the real (albeit indirect) influence of those projects. In addition, we showed that there is limited overlap between the two methods, as reflected in the low number of projects linked to patents that were jointly identified by both approaches. Thus, the two approaches should be considered complementary, providing a broader and more accurate picture (in both the short- and long-term) of how public science funding inspires patent activity.

Finally, we delineated some of the factors behind the likelihood of observing a grant-patent linkage. Specifically, we observed that the matching approach is not particularly appropriate for identifying patents linked to projects with a translational nature. Instead, that method provides stronger results for projects with a higher age, duration, budget, and number and quality of associated publications.

Contributions to prior literature

First, our study contributes to the literature on publication-patent linkages by substantiating the technological impact of public funding for science (Fleming et al., 2019; Li et al., 2017; Jefferson et al., 2018). This field of study is rapidly growing due to the increased availability of large databases of scientific publications, such as Scopus and Lens.org (also available on an open access logics). Scholars can use sophisticated text mining techniques to extract information about funding sources and subsequent patents, and thereby reconstruct linkages. As this type of information becomes increasingly available in a user-friendly format,

we expect that more scientific articles and funding agencies will adopt this approach (Jefferson et al., 2018).

In this context, our results are particularly important for better understanding both the potential and limitations of using the matching method to assess the linkage between scientific research (public or private) and technology development. First, while this method works well for assessing projects' medium- to long-term connections to patented technologies, it can underestimate their short-term effect, especially in relation to the direct contributions of PIs and their teams. In this respect, our data show that the scientific publications that result from publicly funded projects are cited in a very limited share of patents that are developed during the project and reported by the PI to the funding agency as a direct output of the project. The exploratory interviews with PIs and KTOs involved in such grants confirmed that this phenomenon mainly occurs because researchers want to satisfy the novelty requirement and increase their chances of a successful patent process. The interviews also highlighted how PIs and their teams make strategic decisions about the disclosure of scientific results and the timing of publications vis-à-vis patent filing, when they foresee the opportunity and interest for protecting inventions with patents. On the one hand, researchers should be careful when adopting the matching approach, as it can systematically underestimate the direct patenting contribution of investments in public research. On the other hand, our findings call for a deeper exploration of the underlying motivations that lead scientists to cite (or not) their scientific work in patents, along with the effects that might be exerted by organizational influences (such as collaborating with university KTOs) or policy actions (such as the "grace period"[24]).

[24] In some patent systems, the novelty requirement for patents is partly mitigated by allowing researchers to make disclosures of their inventions during a "grace period" (of a few months prior to filing their patent application), and thereby avoid compromising the patentability of their invention.

Second, the matching method can underestimate the influence of projects with a more applied/translational nature vis-à-vis basic research projects, as the former might be less likely to generate scientific publications. This restricts the applicability of the method to specific categories of publicly-funded projects, those with a stronger focus on frontier research outcomes, a valuable lesson for future research in the field of science and innovation funding (Jefferson et al., 2018; Lane, 2009). Third, this approach may perform better in sectors where the innovation cycles are particularly long and complex, such as the life sciences.

Implications for funding agencies and policy-makers

This paper reinforces the importance of promoting technological progress by investing public funds into scientific research (Bercovitz and Feldman, 2006; Broström et al., 2021). We also affirmed that the European Research Council is pivotal in not only promoting excellent science, but also fostering technological and innovation development. Our study presents useful indications for policymakers and managers of funding agencies who want to assess the technological impact of their funded research. First of all, we offer a systematic methodology for collecting, cleaning, and analyzing data on grants, publications and patents. Using our approaches, the administrators of funding agencies can leverage major scientific repositories to determine how scientific projects connect to subsequent technological innovation (via publications). This represents per se a first challenge, considering the issues of data retrieval and cleaning in this area. The second step involves linking such publications to subsequent patents in order to gain a more complete overview of the project's influence. Increasingly, databases of peer-reviewed literature, such as Lens.org and Scopus, are integrating this kind of information in their services. We showed that funding agencies can adopt this approach to capture not only direct patents generated by the research team, even after the end of the reporting period, but also indirect patents generated by other companies or research institutions.

On a related note, our research presents some implications about accountability when generating patents from federally funded research. We observed that direct patents filed by inventors and/or their Host Institution and inspired by public funded research outnumber those that are reported at the Funding agency. The United States previously addressed this issue by introducing accountability safeguards through the 1980 Bay-Dole Act. This legislation requires academic recipients of federal research funding to report not only the existence of federally funded patents, but also information regarding their licensing, assignment and practical utilization (Rai and Sampat, 2012). However, such accountability requirements have not been legally codified in most European countries. Introducing this requirement (e.g., for academic recipients of EU funding when they apply at the EPO for EU-funded patents) could be extremely useful in evaluating the impact of public funding and calibrating the deployment of future funding.

Limitations and future research

Our work highlights several avenues that can be explored by future research. A first limitation of our paper refers to the fact that we were not able to assess the causal impact of receiving ERC funding on the generation of subsequent patented innovations. This important issue went beyond the original intent of our paper, requiring a different methodological approach. In more general terms, the approach we adopt in the paper using patent-to-citations linkages does not allow to capture the evidence of a direct knowledge spillover, but it rather interprets the relations between citing patents and cited scientific work as a proxy of the background scientific knowledge that inspired the technology development, in line with several articles adopting similar methodologies (Azoulay et al., 2007; Marx and Fuegi, 2022; Callaert et al., 2014; Roach and Cohen, 2013).

For a more rigorous understanding on potential impact, thus, future research could gather data and compare different programmes and/or consider scientific publications generated by similar research which did not obtain any type of support. For instance, one possible approach coherent with the data structure presented in this paper would be a matched-paired analysis based on a control group of scientific publications that are homogenous by field, journal and year of development to those generated by research grants, in order to assess their respective likelihood to get cited by future patents. There may also be value in collecting and maintaining information on the ERC applicants whose projects were not selected by the evaluation panels, as such information could be used to develop counterfactual evaluations.

The second interesting opportunity regards a deeper investigation of potential influences exerted by the institutional environment in which the research is conducted. Unfortunately, we had very limited information on the characteristics of the research institutions hosting the projects and the type of stimulus and support they were able to provide also in terms of patenting activity. An interesting dimension to explore in this respect regards the existence of potential differences between universities and public research organizations (PROs). Indeed, in several countries, a large part of fundamental research is conducted by PROs (just to mention a few examples, CNRS or Inria in France; Max Planck or Fraunhofer in Germany; CNR in Italy; in the United States, the Lawrence Berkeley National Lab, Lawrence Livermore National Lab or Sandia National Labs). There has been, however, very limited research focused on comparing universities and PROs in the technology transfer realm (Bozeman and Crow, 1991). Similarly, at least in Europe, legislative reforms and policy interventions to stimulate knowledge transfer have mainly targeted universities, devoting less attention (also in terms of impact assessment) on PROs. However, these two types of institutional actors differ significantly across various dimensions (mission; degree of interdisciplinarity; application orientation; legislative framework for scientists' status), so they may present different rates of patenting activity, both

directly generated and indirectly influenced. The approach presented in this paper could be an effective way to provide new evidence in this respect (for instance, by exploiting the fact that PROs represent an important share of publicly funded research grants).

An additional important aspect that we were not able to directly address in our paper refers to the geographical scope of the technological spillovers stemming from the research projects. For instance, one could argue that research teams that are heterogeneous (for instance in terms of origin of the co-authors of the publications) are more likely to generate a geographically dispersed technological impact, due to the importance of proximity effects and personal interactions for converting scientific results into practical applications. This evidence could have important implications for the design of funding program aimed at stimulating knowledge transfer across countries or regions.

On a similar vein, the context and the approach of our study could open up important avenues for future research on the effects of researchers' mobility (Baruffaldi and Landoni, 2012; Herrera et al., 2010; Veugelers and Van Bouwel., 2015). One of the key peculiarities of the ERC FP7 funding grants was the concept of portability of the grant, meaning that the possibility of transferring the entire project or part of it to a new beneficiary than the initial Host Institution. They represent therefore a very interesting setting for future research to explore whether this type of funding enhances the mobility of researchers and if the results in terms of generation of patents and extent of knowledge spillovers change, by comparing PIs who moved to a different Host Institution to develop the project with those that did not move. This is an interesting and promising area of research as the extant literature on academic mobility shows a positive effect on scientific productivity of researchers. Similarly, we could expect that, by focusing on the process of technological transfer, a similar patter may occur: Researchers moving to a different Host Institution may be less risk-adverse and more able to activate new partnerships and opportunities and, thus, generate more direct and/or indirect

patents. However, it is also known that mobility interrupts consolidated relations and requires researchers to build new contacts with industrial actors, thus making the assessment of the net effect of mobility not straightforward.

References

Auranen, O.; Nieminen, M. (2010) 'University research funding and publication performance—An international comparison'. Research Policy, 39(6), 822-834.

Azoulay, P.; Graff Zivin, J. S.; Li, D.; Sampat, B. N. (2015). 'Public R&D investments and private-sector patenting: evidence from NIH funding rules'. NBER: Working Paper.

Azoulay, P.; Graff Zivin, J. S.; Li, D.; Sampat, B. N.. (2019). 'Public R&D investments and private-sector patenting: evidence from NIH funding rules'. The Review of Economic Studies, 86(1), 117-152.

Baruffaldi, H.S., landoni, P. (2012) 'Return mobility and scientific productivity of researchers working abroad: The role of home country linkages'. Research Policy, 41 (9), 1655-1665.

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

Bercovitz, J., Feldman, M. (2006). Entpreprenerial Universities and Technology Transfer: A Conceptual Framework for Understanding Knowledge-Based Economic Development. Journal of Technology Transfer 31, 175–188.

Broström, A., Buenstorf, G., McKelvey, M. (2021) The knowledge economy, innovation and the new challenges to universities: introduction to the special issue, Innovation, 23:2, 145-162,

Callaert, J.; Pellens, M.; Van Looy, B. (2014) 'Sources of inspiration? Making sense of scientific references in patents'. Scientometrics, 98 (3), 1617-1629.

Cohen, W. M.; Nelson, R. R.; Walsh, J. P. (2002) 'Links and impacts: the influence of public research on industrial R&D'. Management Science, 48 (1), 1-23.

de Rassenfosse; G., Jaffe, A.; Raiteri, E. (2019). 'The procurement of innovation by the US government'. PloS one, 14(8),.1-11.

Dosi, G., Llerena, P., Labini, M. S. (2006). 'The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called 'European Paradox''. Research Policy, 35(10), 1450-1464.

Du, J.; Li, P.; Guo, Q.; Tang, X. (2019) 'Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis'. Journal of Infometrics, 13 (1), 132-148.

Fleming, L., Greene, H., Li, G., Marx, M., Yao, D. (2019). 'Government-funded research increasingly fuels innovation'. Science, 364(6446), 1139-1141.

Geuna, A.; Rossi, F. (2011)' Changes to university IPR regulations in Europe and the impact on academic patenting'. Research Policy, 40 (8), 1068-1076.

Grassano, N.; Rotolo, D.; Hutton, J.; Lang, F.; Hopkins, M.M. (2017) 'Funding data from publication acknowledgments: Coverage, uses, and limitations'. Journal of the Association for Information Science and Technology, 68 (4), 999-1017.

Griliches, Z. (1992). Introduction to output measurement in the service sectors in output measurement in the service sectors. The University of Chicago Press, Chicago, London.

Jaffe, A. B. (1989). 'Real effects of academic research'. The American economic review, 957-970.

Herrera, L., Muñoz-Doyague, M.F., Nieto, M. (2010). ‘Mobility of public researchers, scientific knowledge transfer, and the firm's innovation process’. Journal of Business Research, 63(5): 510-518.

Jaffe, A. (1989). Real effects of academic research. The American Economic Review, 79(5), 957-970.

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

Jefferson, O. A., Jaffe, A., Ashton, D., Warren, B., Koellhofer, D., Dulleck, U., Bilder, G. (2018). 'Mapping the global influence of published research on industry and innovation'. Nature Biotechnology, 36(1), 31-39.

Li, D., Azoulay, P., Sampat, B. N. (2017). 'The applied value of public investments in biomedical research'. Science, 356(6333), 78-81.

Mansfield, E. (1991). 'Academic research and industrial innovation'. Research Policy, 20(1), 1-12.

Munari, F., Toschi, L. (2021). The impact of public funding on science valorisation: an analysis of the ERC Proof-of-Concept Programme, Research Policy, 50(6).

Perkmann, P., Walsh, K. (2009). The two faces of collaboration: impacts of university-industry relations on public research, *Industrial and Corporate Change*, 18(6), 1033–1065

Pisano, G.P. (2010) The evolution of science-based business: Innovating how we innovate. Industrial and Corporate Change, 19 (2), 465–482.

Poege, F.; Harhoff, D.; Gaessler, F.; Baruffaldi, S. (2019). 'Science quality and the value of inventions'. Science Advances, 5(12), 1-8.

Rai, A.K., Sampat, B. N. (2012). 'Accountability in patenting of federally funded research'. Nature biotechnology, 30(10), 953-956.

Rake, B. (2021) 'Do publication activities of academic institutions benefit from formal collaborations with firms? '. Innovation, 23:2, 241-265

Roach, M., Cohen, W.M. (2013). 'Lens or prism? Patent citations as a measure of knowledge flows from public research'. Management Science, 59(2), 504-525.

Tijssen, R.J.W., Buter, R.K., Van Leeuwen, T.N. (2000). 'Technological relevance of science: Validation and analysis of citation linkages between patents and research papers'. Scientometrics, 47, 389–412.

Veugelers, R., Van Bouwel, L. (2015). ‘The effects of international mobility on European researchers: Comparing intra-EU and US Mobility’. Research in Higher Education, 56, 360-377.

Veugelers, R., Wang, J. (2019). 'Scientific novelty and technological impact'. Research Policy, 48 (6), 1362-1372.

Wang, J., Shapira, P. (2011). 'Funding acknowledgement analysis: an enhanced tool to investigate research sponsorship impacts: the case of nanotechnology'. Scientometrics, 87(3), 563-586.

Tables and Figures

Table 1 – List of variables

Variable name	Description
Project with granted patents-publications matches	Dummy equal to 1 if the ERC project leads at least to one granted patent (according to patent-publications-grant matches)
Project declaring granted patents	Dummy equal to 1 if the ERC project declared at least one granted patent (according with the ERC)
Starting Grant	Dummy equal to 1 if the ERC project was a Starting Grant (zero otherwise) (baseline case in the regression analyses)
Advanced Grant	Dummy equal to 1 if the ERC project was an Advanced Grant (zero otherwise)
Consolidator Grant	Dummy equal to 1 if the ERC project was a Consolidator Grant (zero otherwise)
Proof of Concept Grant	Dummy equal to 1 if the ERC project was a Proof of Concept Grant (zero otherwise)
Life Science	Dummy equal to 1 if the ERC project was in the Life Science field (zero otherwise)
Project age	Age in years of the ERC project assessed as difference between 2020 and the starting year of the project
Number of partner institutions	Number of partner institutions collaborating in the ERC project
Project budget (million euro)	Amount of project budget received by the ERC (in millions of Euros)
Project duration	Duration of the ERC project in years assessed as difference between the ending year and starting year of the project
Total publications	Number of publications generated in CORDIS and Scopus by the ERC project up to 2020
Total citations	Number of citations generated in Scopus by the publications associated to the ERC project up to 2020
Eastern Europe	Dummy equal to 1 if the ERC project host institution is located in Eastern Europe (zero otherwise)
Northern Europe	Dummy equal to 1 if the ERC project host institution is located in Northern Europe (zero otherwise)
Southern Europe	Dummy equal to 1 if the ERC project host institution is located in Southern Europe (zero otherwise)
Western Europe	Dummy equal to 1 if the ERC project host institution is located in Western Europe (zero otherwise)

Table 2 - Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Project with granted patents-publications matches	3697	0.29	0.45	0	1
Project declaring granted patents	3697	0.04	0.20	0	1
Starting Grant	3697	0.51	0.50	0	1
Advanced Grant	3697	0.38	0.48	0	1
Consolidator Grant	3697	0.07	0.25	0	1
Proof of Concept Grant	3697	0.04	0.20	0	1
Life Science	3697	0.44	0.50	0	1
Project age (in years)	3697	8.40	1.83	5	12
Number of partners institutions	3697	0.21	0.51	0	5
Project budget (million euro)	3697	1.69	0.61	0.11	3.50
Project duration (in years)	3697	4.84	1.01	0	8
Total publications	3697	27.6	26.41	0	189
Total citations	3697	912.13	1,799.3	0	43,927
Eastern Europe	3697	0.02	0.13	0	1
Northern Europe	3697	0.29	0.45	0	1
Southern Europe	3697	0.18	0.38	0	1
Western Europe	3697	0.52	0.50	0	1

Table 3 - Correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) Project with granted patent-publications matches	1.00																
(2) Project declaring granted patents at ERC	0.16*	1.00															
(3) Starting Grant	0.03*	-0.01	1.00														
(4) Advanced Grant	0.08*	0.04*	-0.80*	1.00													
(5) Consolidator Grant	-0.12*	-0.02	-0.28*	-0.21*	1.00												
(6) Proof of Concept Grant	-0.13*	-0.04*	-0.22*	-0.16*	-0.06*	1.00											
(7) Life Science	0.00	-0.03	0.00	0.01	0.01	-0.04*	1.00										
(8) Project age	0.31*	0.04*	0.23*	0.02	-0.36*	-0.17*	-0.04*	1.00									
(9) Number of partners	-0.01	0.01	-0.00	0.02	-0.01	-0.04*	-0.00	0.01	1.00								
(10) Project budget	0.12*	0.07*	-0.50*	0.68*	0.09*	-0.52*	0.16*	-0.11*	0.06*	1.00							
(11) Project duration	0.09*	0.03	0.15*	0.16*	0.04*	-0.81*	0.05*	0.20*	0.07*	0.48*	1.00						
(12) Total publications	0.20*	0.05*	-0.08*	0.21*	-0.07*	-0.21*	-0.34*	0.19*	0.05*	0.13*	0.18*	1.00					
(13) Total citations	0.30*	0.04*	-0.01	0.09*	-0.08*	-0.10*	-0.06*	0.25*	0.01	0.09*	0.08*	0.44*	1.00				
(14) Eastern Europe	0.02	0.03*	0.00	-0.00	-0.00	0.00	-0.04*	0.02	0.04*	-0.05*	-0.01	0.02	-0.02	1.00			
(15) Northern Europe	-0.02	-0.04*	-0.03	0.05*	-0.03*	-0.00	0.01	-0.01	0.02	0.04*	0.02	0.01	0.02	-0.09*	1.00		
(16) Southern Europe	0.02	0.05*	0.03	-0.05*	0.03	0.01	0.01	0.03	0.01	-0.07*	0.01	0.02	0.00	-0.06*	-0.30*	1.00	
(17) Western Europe	-0.01	-0.01	0.01	-0.01	0.01	-0.01	-0.00	-0.02	-0.04*	0.03	-0.03	-0.03*	-0.01	-0.14*	-0.66*	-0.48*	1.00

* p-level < .05

Table 4 - Distribution of patent outcomes by ERC sector and grant

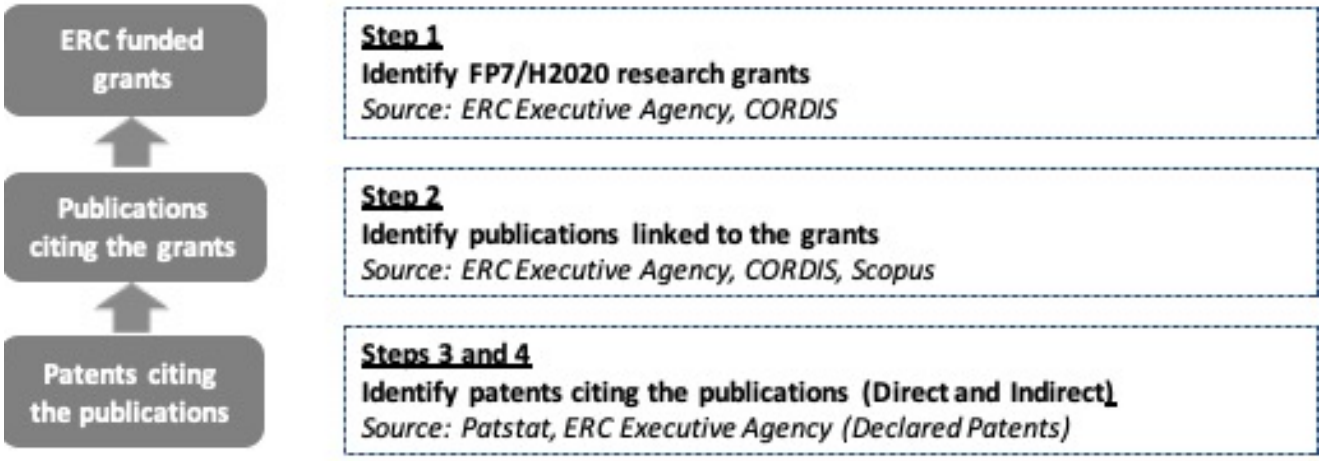
		Total Projects	Number of projects with publications cited by subsequent granted patents	Share of projects with publications cited by subsequent patents	Number of projects declaring a patent in the final reports	Share (%) of projects declaring a patent in the final reports
Total	**Advanced Grant**	1,397	473	33.86%	70	5.01%
	Consolidator Grant	254	24	9.45%	7	2.76%
	Starting Grant	1,889	578	30.60%	71	3.76%
	Total	3,697	1,076	29.11%	148	4.00%
Life Science	**Advanced Grant**	627	228	36.36%	32	5.10%
	Consolidator Grant	115	12	10.43%	4	3.48%
	Starting Grant	837	238	28.43%	19	2.27%
	Total	1,632	478	29.29%	55	3.37%
Physical Engineering	**Advanced Grant**	770	245	31.82%	38	4.94%
	Consolidator Grant	139	12	8.63%	3	2.16%
	Starting Grant	1052	340	32.32%	52	4.94%
	Total	2,065	598	28.96%	93	4.50%

Table 5 – Regression analyses

	(1)	(2)	(3)	(4)
	Project with publications cited by subsequent granted patents		**Project declaring a patent in the final report**	
Project-level characteristics				
Advanced Grant	-0.334***	-0.472***	-0.247*	-0.264**
	(0.076)	(0.079)	(0.127)	(0.129)
Consolidator Grant	-0.361***	-0.392***	-0.209	-0.216
	(0.125)	(0.131)	(0.191)	(0.192)
Proof of Concept Grant	-1.532***	-0.773*		
	(0.403)	(0.415)		
Life Sciences	-0.058	0.167***	-0.205**	-0.166*
	(0.048)	(0.055)	(0.080)	(0.086)
Project age (in years)	0.261***	0.206***	0.060**	0.050**
	(0.014)	(0.015)	(0.023)	(0.024)
Number of partner institutions	-0.119	-0.158**	-0.050	-0.056
	(0.077)	(0.079)	(0.126)	(0.126)
Project budget (million euros)	0.642***	0.639***	0.516***	0.514***
	(0.071)	(0.074)	(0.118)	(0.118)
Project duration (in years)	-0.191***	-0.188***	-0.090	-0.089
	(0.038)	(0.040)	(0.063)	(0.064)
Scientific production of the project				
Total publications		0.266***		0.053
		(0.035)		(0.047)
Total citations		0.000***		0.000
		(0.000)		(0.000)
Constant	-2.782***	-3.405***	-2.706***	-2.811***
	(0.241)	(0.269)	(0.398)	(0.419)
Observations	3,697	3,697	3,540	3,540
Pseudo R^2	0.122	0.193	0.038	0.041
European region dummies	YES	YES	YES	YES

Standard errors in parenthesis
**** p<0.01, ** p<0.05, * p<0.1*

Figure 1 – Methodological approach



Online Supplementary Materials for the article

"Public research funding and science-based innovation: an analysis of ERC research grants, publications and patents"

Online Tables

Online Table 1A - A summary of research papers linking publicly supported research grants to patents (via scientific publications cited in such patents)

Paper	Institution	Sector	Scientific Publication Database	Patent Database	Matching approach between patents and publications	Results
Azoulay et al. 2015	U.S. National Institutes of Health grants (NIH)	Biomedical Research	PubMed	USPTO	Direct patents: Match between NPL mentioned in patents at USPTO with metadata in PubMed using publication's acknowledgment. Indirect patents: Match between NPL mentioned in patents at USPTO with metadata in Pubmed using "intellectual similar" publication's acknowledgment. Keep match when at least 4 fields are the same.	44% of NIH funding are directly related to life-sciences patents at USPTO database. 50% of NIH grants are directly linked to high value patents of advanced drug candidates. 84% of NIH funding are indirectly related to patents. 43% of the NIH grants produced a publication directly cited by patents.
Li et al., 2017	U.S. National Institutes of Health grants (NIH)	Life Science	PubMed	USPTO	Match between NPL mentioned in patents at USPTO with metadata in PubMed. Keep match when at least 4 fields are the same	10% of NIH grants connected to a patent directly, 30% generate articles cited by patents.
De Rassenfosse et al., 2019	U.S. government funding	Not specified	Web of Science (WoS)	USPTO and Patsat	From PATSTAT and USPTO database collect patent-level information. Then query WoS API for relevant procurement contract and grant identifiers in the acknowledgement section of scientific publications published between 2000 and 2015	1.5% of the R&D contracts are connected to a granted patent, these contracts represent 36% of the overall value. 76% of the grants related to patents are linked to at least 1 publication. 16% of the procurement contracts related to patents are related to at last 1 publication

Online Table 1B: A summary of research papers linking publicly supported research grants to patents (via scientific publications cited in such patents)

Paper	**Institution Investigated**	**Sector**	**Scientific Publication Database**	**Patent Database**	**Matching approach between patents and publications**	**Results**
Fleming et al., 2019	U.S. Federally supported research	Not specified	Web of Science (WoS)	USPTO	Match between NPL mentioned in patents and entries on WoS based on 4 main metadata. If match is found, then check acknowledgement in NPL	13,75% of the patents generated from 1926 and 2017 are connected to a Federal research project. Precision of patent-to-article algorithm is 100% (all correct items out of 100 random matches).
Marx and Fuegi, 2019	Not Specified (U.S. and non-US patents	Not specified	Microsoft Academic Graph	USPTO	First loose match between patent information and main publication metadata, then re-weighting of the match list based on frequency of key terms in title, author names, journal name	Only 4% of Patent Citations to Science (PCS) are added by patent examiners. 17.6% of patents granted since 1947 contains at least one publication on their front page. 1.5% of all papers are cited by the patent front page. Evaluation carried out at different confidence levels. With confidence level 4 (best result), Precision=99.4% and Recall=92.9%
Jefferson et al., 2018	Not specified	Not specified	Pubmed and Crossref	EPO's citation database	Match between NPL at EPO's citation database and acknowledgment of publications.	11.8 million scholarly outputs from 200 Global Research Institutions extracted from the 1980 to 2015 time period analyzed, 10%, on average, were matched in the Lens citation database. These were cited in 690,000 simple patent families

Online Table 2: Type of ERC grants and requirements (as of October 2020)

Type of Grant	**Researcher**	**Max Budget**	Max duration	**Objectives of the Grant**
Starting Grant	2-7 years after Ph.D.	€ 1,5 million	up to 5 years	Early-career scientists with an excellent resumè and want to work independently.
Consolidator Grant	7-12 years after Ph.D.	€ 2 million	up to 5 years	Scientists seeking to consolidate their career and build a research team
Proof of Concept	Principal Investigator of a preview ERC project	€ 150.000	up to 18 months	For researchers with a previous ERC grant to explore the commercial or societal results
Advanced Grant	Active researchers in the last ten years	€ 2,5 million	up to 5 years	For established scientist looking for a long-term funding to develop novel and risky projects
Synergy Grant	Group of 2-4 Principal Investigators	€ 10 million	up to 6 years	Enhance collaborative projects to work on ambitious projects.

Online Appendix A: Identifying FP7 ERC projects

In the first step of our data collection process, we identified ERC supported projects in the framework programs FP7 and H2020. Using data from the European Union Open Data Portal, we gathered during this step 4,563 FP7 projects and 6,379 H2020 projects, building a dataset with 10,942 projects. As we explain in the Methodology section of the paper, we focused the analysis of this study on FP7 projects, since they were already completed at the time of our study, and to have a longer time span to observe the impact on publications and subsequent patents. Those were distributed through the five grant categories, as listed above in Table A1.

Table A: Distribution of ERC FP7 Projects

Type of ERC Grant	**Total of Projects**	**%**
Starting Grant	2,332	51.10%
Advanced Grant	1,709	37.45%
Consolidator Grant	313	6.86%
Proof of Concept	177	3.88%
Synergy Grant	24	0.53%
Support	8	0.18%
Total	4,563	

According to the E.U. Open Data, from the FP/ projects identified, eight are Support (6) projects, which refers to ERC's internal projects that provides administrative and operational support to scientific councils. We eliminated them from our database since we wanted to focus on research projects. We also removed a Proof of Concept project for which we were not able to collect specific information. In this way, our database has 4,563 projects, distributed in the following scientific domains identified by the ERC: 35.85% of projects were from the Life

Sciences domain (1,633); 45.38% of projects from the Physical and Engineering field (2,067); 18.09% of projects were from the Social Sciences and Humanities domain (824). Note that we assigned Proof-of-Concept projects to the scientific domain of their corresponding previous Frontier Research project (Starting Grant, Advanced Grant, Consolidator Grant), identified by the name of the same Principal Investigator. Indeed, PoC projects are not typically assigned to a specific scientific domain by CORDIS.

As highlighted on Methodology, our final sample included only FP7 projects from the Life Sciences and Physical and Engineering domain, thus corresponding to a total of 3,697 projects. It is necessary to underline that we did not consider Synergy Grants' projects in our final sample, given their limited number (24) and their multidisciplinary nature (so it was not possible to assign them to a specific domain).

We then collected specific information about each of these projects by consulting the website of CORDIS[25] on the following variables:

- Number of the Project;
- Type of ERC Grant;
- Science field;
- Year of call;
- Start Year;
- End Year;
- Budget;
- Host Institution;

[25] Information available at https://cordis.europa.eu/ . We also integrated the data available on the CORDIS website with data available on the European Research Council website

- Country of the Host Institution;
- Partner Institutions;
- Principal Investigator's first and last name;

Online Appendix B: Publications Dataset

In order to identify the scientific publications resulting from the selected projects, we resorted to CORDIS and Scopus.

Publications from CORDIS

On CORDIS, each ERC-project has a dedicated page[26]. We retrieved the list of publications reported in the section "Results". In total the list includes 96,886 publications for ERC FP7 projects. Notably, on CORDIS, the information on publications for later projects (e.g., H2020) is far more structured and complete than the one available for FP7, where authors could insert information in a freeway into the same field. The information for the ERC-related publications reported by CORDIS relied on OpenAire data (https://www.openaire.eu/).

Publications from Scopus

Scopus' research looked for publications that declared on their "Acknowledgments" section that the study was funded by the ERC funding program, using the project number we got from the first step (see Appendix A). Scopus analyses the funding sources cited in the scientific publications' acknowledgments and inserts them in a dedicated metadata field. We downloaded all the publications citing an ERC grant number reported in the dataset identified in the first step. Noteworthy, the information available for each publication retrieved from Scopus is highly homogeneous and detailed. It is worth mentioning that CORDIS is specifically targeting EU-funded outcomes of any type (indexed and non-indexed journals, reports, books,

[26] In CORDIS, the scientific results of each project are generally those declared by the principal investigators, in particular in the case of H2020 projects. In the case of the FP7 program, CORDIS does not rely on the self declaration but rather use data provided by the Portal OpenAIRE, a database aiming to provide an open communication between scholars. OpenAIRE is an EC-supported initiative to foster Open Science in Europe in order to accelerate research and boost innovation.

book chapters, and working papers). On the other hand, Scopus covers scientific publications and books and may not consider different outcomes.

From Scopus, we downloaded 46,061 publications for ERC FP7 projects. Noteworthy, the information available for each publication retrieved from Scopus is highly homogeneous and detailed.

Merge between Scopus and CORDIS publications

The combination of the two datasets contained a partial overlap. We want to identify duplicates from CORDIS and Scopus, i.e., the same publication on both websites, and merge it into a single entry, keeping both websites' metadata. We matched the publications from each source based on having the same title and/or DOI. To reduce the number of false negatives, seeking to increase these matches' quality, we normalized the title by turning text to lowercase and removing accents and non-alphanumeric characters (e.g., apostrophes and dashes). To avoid multiple matches generated by some short and generic publications' titles, we checked manually those papers that matched only the title (e.g., "Alzheimer's disease").

After the merge process, the dataset contains 122,817 unique publications for ERC FP7 projects from LS and PE sectors. From the match between CORDIS and Scopus results, we found that 22,130 of the publications are present in both sources. This overlap shows that about half of publications listed as ERC on Scopus are also on CORDIS (48%). Conversely, only about 22% of CORDIS's publications was in Scopus or at least listed in Scopus as a result of an "ERC" project.

To increase the information available for each publication (information in CORDIS was often not complete), we then searched in Scopus the publications found only in CORDIS. To ensure that the publication mentioned in CORDIS corresponded to the one in Scopus, we checked the DOI. We also looked for publications matching the title and containing the

Principal Investigator (PI) as one of the authors' ERC project. With this step we identified 52,436 publications in Scopus.

The difference between the total of publications between the two sources is interesting for several factors. The number of identified ERC-publications is consistently lower in Scopus compared to CORDIS. However, as Scopus distinguishes ERC-publications through the acknowledgment section, we know that this section became mandatory only for more recent ERC-projects. In this sense, we can conclude that many ERC-publications missed acknowledging the ERC funding, even though they are present on the Scopus website.

After the merge, the dataset counts with 120,800 unique publications. However, as discussed above, given that in the CORDIS' database detailed information on the publication was often lacking, especially for FP7 projects, for our analyses in the end we considered only the subset of publications included in Scopus' dataset, totaling to 86,630 unique publications generated from ERC FP7 projects from the L.S. and P.E. sectors.

Online Appendix C: Patents Dataset

With the list of publications per project, the next step sought to identify which patents cited those publications during their patenting process using the Patstat database. To do that we used Patstat 2019 Autumn Edition. The link between publications and patents involved two different actions.

Linking ERC-publications with Non-Patent-Literature (NPL) in Patstat

The first action was to match the publications gathered with the Non-Patent Literature field of Patstat. To present the state of the art and knowledge present on the background of an invention, an inventor can cite other patents and publications (scientific or others). Examiners also provide this information when evaluating the degree of novelty during the granting process on search reports. The publications' citations are present on a specific table of Patstat, table 214 npl_publn, called Non-Patent Literature (NPL).

This field on Patstat has over 44 million entries, but they are not on the same configuration. Most of them (about 35 million) is unstructured, i.e., a single free text filed keeps all the publication information. And about 9 million are structured entries, which means that there are different fields for each type of data (e.g., title, authors, DOI, year, etc.). To extract the output of ERC-projects related to patents, we matched the 44m NPL papers in Patstat with the 120,000 papers from FP7 projects. The match between publications and patents considered three criteria: the same first author (surname), same title, and same DOI. For NPL instances with structured metadata, we match ERC publications' information only with the related Patstat field. In this step, we considered only publications resulting from exact matches, i.e., when the entire surname, title, or DOI was the same. When dealing with unstructured metadata, we combined the author, title, and DOI with all the information available to identify them regardless of their position.

At this point, it is necessary to underline that before performing the match, we normalized data of authors and titles, i.e., we set all the text to lowercase, removed accents on letters and non-letter characters (e.g., apostrophes and dashes). The data's normalization allowed us to overcome issues when matching surnames, minimizing the risk of accents or formulas on titles that could inhibit the match for being written slightly differently. For each kind of data structure, we dealt with different procedures.

For structured metadata, we matched the information from ERC publications only with the related Patstat field (e.g., table TLS214_NPL_PUBLN, fields: npl_author, npl_title, npl_doi). In this step, we considered only publications resulting from exact matches, i.e., when the entire surname, title, or DOI is the same. We did not consider combinations due to substring matching (e.g., "organic compounds" does not match with "macromolecular organic compounds"). On the other hand, to deal with unstructured metadata, we matched author, title, and DOI with all the information available to identify them regardless of their position.

As a result of this analysis, a publication can match one or more of the following criteria: title, DOI, and author. A higher number of hits means a more confident match (e.g., if a publication matched title, DOI, and author is the best result we can obtain). In total, we have seven possible combinations: matching only DOI; matching only title; matching only author; matching DOI and author; matching title and author; matching DOI and title; matching author, title, and DOI. A priori, we discarded matches based only on the first author surname, given the very high number of results produced and the low significance of homonymy instances. We than considered acceptable the matches with two or more hits, and we manually controlled the publications that matched only title or DOI.

After the manual cleaning, we identified 14,483 NPLs that matched with 5,230 publications. Those publications are linked to 1,545 FP7 projects. The project with the highest number of matches presented 263 correspondences from 23 different publications.

If we consider only the final sample of ERC FP7 projects from L.S. and P.E. domains, we found 1,547 projects whose scientific publications were cited by subsequent patents in the NPL sections. The restricted match we used aimed to avoid false positive matches. After the consolidation of the database, we randomly selected 100 patents to verify if there were false positives. In particular, we manually accessed the selected patents through Espacenet[27] and searched for the publication on the citation section of the patents. Then, we searched for the publications on Scopus to verify if they were outcomes of the referred project. We did not find any false positive on this sample.

From NPL to Patents

We used as a patent identification the application number of the patent, that is, the number that identifies the beginning of the procedure to a specific office. To pinpoint the patents linked to these publications was necessary to connect the NPL identifications with other tables inside Patstat. In this way, for each publication found in the previous match, we identified the patent's application number to which it is linked. We were able to identify 12,918 unique patents linked to 1,547 FP7 LS and PE projects. For each patent relying on ERC-related publicatons, we collected from Patstat the following additional information: application identification, inventor and applicant identification, if it is a granted patent, patent citations made, NPL citations made, citations by other patents, patent family identification, patent family authorities, priority ear, application year, and International Patent Classification (IPC).

[27]Espacenet is an open database from European Patent Office that provides access to patents from 1782 to today. The database is available on (https://worldwide.espacenet.com/)

Online Appendix D: T-Test Linked and Declared Patents

To better characterize our sample, we use a t-test to determine whether projects cited by patents are different from projects not cited. Also, we provide a similar comparison for projects with declared patents vis-à-vis projects not declaring this type of output (see Online Appendix E). As can be seen from the difference test (both t-test and its p-value) of this Appendix, several variables are significantly different between the two groups.

Table D: T-test on proportion difference of the explanatory variables by project with/without linked granted patents

	Project *with* publications cited by subsequent granted patents			**Project *without* publications cited by subsequent granted patents**			**Difference-test**	
	Obs.	**Mean**	**S.E.**	**Obs.**	**Mean**	**S.E.**	**t-test**	**Sign.**
Starting Grant	1076	0.537	0.015	2621	0.500	0.010	-2.043	**
Advanced Grant	1076	0.440	0.015	2621	0.352	0.009	-4.959	***
Consolidator Grant	1076	0.022	0.005	2621	0.088	0.006	7.146	***
Proof of Concept Grant	1076	0.001	0.001	2621	0.060	0.005	8.025	***
Life Science	1076	0.444	0.015	2621	0.440	0.010	-0.220	n.s.
Project age (years)	1076	9.295	0.052	2621	8.030	0.034	-20.104	***
Number of partners institution	1076	0.204	0.015	2621	0.218	0.010	0.775	n.s.
Project budget (million euros)	1076	1.799	1.569	2621	1.637	1.255	-7.322	***
Project duration (years)	1076	4.980	0.019	2621	4.780	0.022	-5.521	***
Total publications	1076	35.696	0.796	2621	24.280	0.504	-12.174	***
Total citations	1076	1746.125	86.171	2621	569.747	18.385	-18.908	***
Eastern Europe	1076	0.021	0.004	2621	0.017	0.003	-0.950	n.s.
Northern Europe	1076	0.277	0.014	2621	0.293	0.009	0.980	n.s.
Southern Europe	1076	0.191	0.012	2621	0.173	0.007	-1.248	n.s.
Western Europe	1076	0.511	0.015	2621	0.517	0.010	0.322	n.s.

*** p<0.01, ** p<0.05, * p<0.1

Online Appendix E: T-Test Declared Patents

Table E: T-test on proportion difference of the explanatory variables by project with/without declared granted patents by the ERC

	Project *declaring a patent* in the final report			**Project *NOT declaring* a patent in the final report**			**Difference-test**	
	Obs.	**Mean**	**S.E.**	**Obs.**	**Mean**	**S.E.**	**t-test**	**p-value**
Starting Grant	148	0.480	0.041	3392	0.536	0.009	1.342	*
Advanced Grant	148	0.473	0.041	3392	0.391	0.008	-1.992	**
Consolidator Grant	148	0.047	0.017	3392	0.073	0.004	1.178	n.s.
Life Science	148	0.372	0.040	3392	0.449	0.009	1.861	*
Project age (in years)	148	8.764	0.149	3392	8.449	0.032	-2.042	**
Number of partner institutions	148	0.230	0.047	3392	0.217	0.009	-0.294	n.s.
Project budget (million euros)	148	1.889	0.041	3392	1.746	0.009	-3.203	***
Project duration (in years)	148	4.980	0.056	3392	5.011	0.010	0.611	n.s.
Total publications	148	33.574	2.251	3392	28.551	0.452	-2.269	**
Total citations	148	1303.878	30.884	3392	935.579	30.907	-2.400	**
Eastern Europe	148	0.041	0.016	3392	0.017	0.002	-2.095	**
Northern Europe	148	0.196	0.033	3392	0.292	0.0.008	2.537	**
Southern Europe	148	0.270	0.037	3392	0.173	0.007	-3.010	**
Western Europe	148	0.493	0.041	3392	0.517	0.009	0.561	n.s.

*** p<0.01, ** p<0.05, * p<0.1