Technological Forecasting & Social Change 184 (2022) 122009

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

# Predictors of applying for and winning an ERC Proof-of-Concept grant: An automated machine learning model

Marco Seeber [a,*], Ilan Alon [b,c], David G. Pina [d], Fredrik Niclas Piro [e], Michele Seeber [f]

[a] University of Agder, Department of Political Science and Management, Kristiansand, Norway
[b] University of Ariel, Department of Economics and Business Administration, Ariel, Israel
[c] University of Agder, School of Business and Law, Post Box 422, Kristiansand, Norway
[d] European Research Executive Agency, European Commission, Brussels, Belgium
[e] Nordic Institute for Studies in Innovation, Education and Research (NIFU), Oslo, Norway
[f] University of Modena and Reggio Emilia, Department of Life Sciences, Modena, Italy

ARTICLE INFO

*Keywords:*
Research proposals evaluation
Automated machine learning
ERC
PoC
Research valorization
Research funding
Likelihood to apply
Artificial intelligence

ABSTRACT

Research often fails to be translated into applications because of lack of financial support. The Proof of Concept (PoC) funding scheme from the European Research Council (ERC) supports the early stages of the valorization process of the research conducted by its grantees. This article explores the factors that predict who will apply for ERC grants and which grant proposals will prove successful. By combining information from two datasets of 10,074 ERC grants (representing 8361 individual grantees) and 2186 PoC proposals, and using automated machine learning, we can identify the main predictors of the propensity to apply and to win. Doing so fills a void in the literature on likelihood to apply. The results reveal major differences between potential and actual beneficiaries, due to decisions about applying for a grant and evaluations of the proposals. The decision to apply is affected by the interaction between the characteristics of the PoC funding scheme, the ERC grantee, and his/her environment. Grantees in countries that invest little in innovation, with low cost of personnel, and strong collaboration in innovation are more likely to apply. Male grantees are more likely to apply but have similar chances of winning as women.

## 1. Introduction

Research activity is expected to contribute to economic and social development. However, translating scientific knowledge into social and commercial applications is quite challenging and uncertain, especially in the case of potentially breakthrough applications (Vilkkumaa et al., 2015). Private investors tend to be reluctant to invest in such a risky endeavor (Bruneel et al., 2010), leading to a notorious "funding gap" in the early stages of the valorization process (e.g., Lockett and Wright, 2005; Rasmussen and Sørheim, 2012). Consequently, many potentially innovative ideas fail to be translated into useful applications (Auerswald and Branscomb, 2003), especially because of a lack of financial support during the concept stage (García-Quevedo et al., 2018). To address this issue, universities, governments, and public agencies have recently introduced funding schemes to support the valorization of research activity (Bradley et al., 2013; Munari et al., 2018; Rasmussen, 2008).

A notable example is the Proof of Concept (PoC) funding scheme of the European Research Council (ERC), the European Union's flagship program supporting individual, "frontier" research. The PoC was established in 2011 to promote the early stages of the valorization process of the research conducted by ERC grantees. Munari and Toschi (2021) recently showcased the effectiveness of receiving a PoC grant, particularly for early career researchers.

However, studying the effects of receiving a grant provides only a partial view of the ability of a funding scheme to achieve its mission. It is comparable to assessing the success of a humanitarian food program by looking only at the effects on those who received food. There is a potential "iceberg" effect: the visible part, that is, the beneficiaries, can be just a fraction of those truly needing or asking for help. In order to investigate the ability of a research funding scheme to achieve its objectives, we should also examine the selection phase (Enger and Castellacci, 2016; European Commission, 2015; Hoenig, 2017; Neufeld

* Corresponding author.
*E-mail addresses:* marco.seeber@uia.no (M. Seeber), ilanal@ariel.ac.il, ilan.alon@uia.no (I. Alon), david.pina@ec.europa.eu (D.G. Pina), fredrik.piro@nifu.no (F.N. Piro), michele.seeber@unimore.it (M. Seeber).

https://doi.org/10.1016/j.techfore.2022.122009
Received 26 October 2021; Received in revised form 28 June 2022; Accepted 29 August 2022
Available online 8 September 2022

et al., 2013), namely, who applies from the pool of potential beneficiaries and which proposals win.

While several studies have explored the factors predicting a proposal's success, research on what factors predict the likelihood of applying for a grant is almost absent. This is a glaring gap because the decision to apply for a grant does not depend exclusively on a scientist's personal interest in valorization or the potential of a research finding to be translated into a commercial application. Other factors such as the characteristics of the grant, its funding amount, duration, the context in which a scientist works, as well as the characteristics of the evaluation process may subtly affect who applies, and hence influence the ability of the funding scheme to achieve its mission.

The main objective of this article is therefore to fill this gap by exploring what factors predict the likelihood of potential beneficiaries, i.e., winners of an ERC grant, applying for a valorization grant scheme, i.e., the ERC's PoC grant, and what factors predict which proposals will be successful. To do so, we combine information from two datasets: i) a dataset including information on 10,074 ERC grants and the 8361 individual ERC grantees, namely, the potential beneficiaries; ii) a dataset on 2186 PoC proposals, which have been developed by ERC grantees. We then use a machine learning approach to identify the characteristics of an ERC grantee and a PoC proposal that predict the likelihood of applying for and winning a PoC grant. To our knowledge, this is the first study to explore the likelihood of applying in addition to the likelihood of winning one.

The article is also innovative with regard to its methodological approach. It is the first study of the grant selection process to use automated machine learning (AML). AML allows researchers to explore complexity using big data and to validate theories and causal patterns derived from the data. AML is a new predictive analytics technology that is increasingly used in industry and academia to find the optimal model with the fewest predictive errors across multiple segments of the data to understand and forecast social and physical phenomena (Truong et al., 2019). While its application in the social sciences is still limited, recent research suggests that it may be superior to the traditional regression methods based on full sample estimations (Doornenbal et al., 2021). In AML, computation processes automatically "learn" patterns and improve predictions (Kuhn and Johnson, 2013), a process attributed to limited artificial intelligence (AI). AML models have been utilized in various fields, including marketing (Alon et al., 2001), human resources and leadership (Doornenbal et al., 2021), strategy (Alon et al., 2015), finance (Munim et al., 2019), and logistics (Munim and Schramm, 2021). AML is seen as an abductive empirical technique. Hypotheses are derived by assessing observations in light of theory and then testing, validating, and cross validating the data on unseen data segments (Doornenbal et al., 2021; Mantere and Ketokivi, 2013).

The following section discusses the challenges for translational research and the emergence of innovative funding tools, as well as the existing knowledge on the factors that affect the likelihood of a researcher applying for a grant and of a proposal succeeding. We then present the data and the methods of analysis, followed by the results section. The final section discusses the article's main findings and their implications for scholars and policy makers interested in research grant schemes, evaluation procedures, and research valorization instruments.

## 2. Research valorization and project funding

### 2.1. The European Union's pursuit of excellence and innovation

One of the objectives of publicly funded research is to create economic and societal wealth that benefits citizens. For this to happen, the outputs originating from research activities should ideally be translated into products, services, and innovations. Europe has had some difficulties in achieving this objective. The European Commission has acknowledged this problem (European Commission, 1994) and raised it to the top of the political agenda, calling it the "European paradox" (European Commission, 1995). The term refers to the belief that European universities and research centers are among the world's top-level research producers in terms of the number of scientific publications but is lagging behind US universities when it comes to translating these results into applications or innovations. The existence of such a paradox has been the subject of debate. Some scholars have argued Europe is also lagging behind the US in terms of scientific quality and impact, given its smaller share of highly cited publications (Albarrán et al., 2010; Dosi et al., 2006; King, 2004; Rodríguez-Navarro and Narin, 2018).

Irrespective of whether the European paradox truly exists, EU political leaders have a long-term commitment to excellent scientific research. Beginning with the Lisbon Strategy, they decided to transform the EU into the most competitive and knowledge-based economy (European Council, 2000). The creation of the ERC was the main method for addressing this challenge and became a reality with the launch of the Seventh Framework Program (FP7) in 2007.

However, ensuring that the results of research are translated into applications or commercialized into new products and services is challenging. The so-called "Valley of Death" corresponds to the phase in the innovation value chain where funds are lacking (Munari and Toschi, 2021), namely the period related to the testing, validation or demonstration of an idea or invention. These activities in the early stage of the valorization process struggle to attract the interest of private funding investments such as seed investors or business angels (Rasmussen and Sørheim, 2012). The ERC's PoC funding instrument was introduced in 2011 to tackle this problem, and promote the translation of ideas, discoveries and inventions emerging specifically from ERC fundamental "frontier" research projects. With this funding instrument, ERC grantees who receive the grants are given the chance to translate their ideas into social or technological innovations that can ultimately bring societal and economic benefits to Europe and its citizens (Munari and Toschi, 2021).

### 2.2. Factors predicting the likelihood of applying for a research grant

There are very few studies exploring what factors and traits predict the likelihood of applying for a research grant. In a study focusing on research organizations in Norway, Enger and Castellacci (2016) found that prior participation in the EU's Framework Programs and the existence of complementary national funding scheme predict the likelihood of applying. In a similar vein, Bol et al. (2018) reported that applicants to the Netherlands Organization of Scientific Research (NWO) who fail to win one grant apply less often in the future. Enger (2018) suggested that participation in Horizon 2020 is positively related to the existence of closed clubs composed of European universities that have a strong and influential network in collaborative research.

The likelihood of applying for a research grant also depends on the characteristics of the funding scheme. Neufeld et al. (2013), for example, observed that applicants for ERC Starting Grants (StGs) have an above-average scientific output and impact. They explained this outcome as the result of a self-selection process, whereby researchers recognize the highly selective and demanding nature of the ERC StG funding scheme. Therefore, only those who have a strong scientific track record apply for it.

The PoC funding scheme displays some specific features that are expected to affect the likelihood of an ERC grantee applying for it. These features include the fact that it supports translational research in its early phases and provides financial support of 150,000 euros.

Studies on academic entrepreneurship have identified some traits that are predictive of the level of engagement with translational research, engagement with industry, and/or commercialization (e.g., D'Este and Patel, 2007; Lam, 2011). Two individual traits that are consistently predictive of translational research are scientific productivity and gender (Perkmann et al., 2013). Scientifically prolific researchers are much more likely to engage with industry and be involved in commercialization and patenting (Azoulay et al., 2007; Ding and Stuart, 2006; Gulbrandsen and Smeby, 2005; Haeussler and Colyvas, 2011; Stephan et al., 2007). In this respect, PoC eligible applicants are researchers who have succeeded in the very competitive ERC grant selection process and among a pool of potential applicants who already display an above-average scientific output and impact (Neufeld et al., 2013). Research on academic entrepreneurship has also revealed that – ceteris paribus – female scientists are generally less likely than their male peers to participate in commercial and technology transfer activities, due to different attitudes, opportunity structures, and structural barriers (e.g., Azoulay et al., 2007; Giuri et al., 2019; Haeussler and Colyvas, 2011; Meng, 2016; Murray and Graham, 2007; Tartari and Salter, 2015). Furthermore, both worldwide and in Europe, female researchers are under-represented among inventors, submitting far fewer patent applications than men (European Commission, 2021, p. 248). Past studies have explored whether gender predicts success in receiving grants. However, there are no studies regarding whether gender predicts the likelihood of applying for a grant. Hence, in line with findings from research on academic engagement with industry and commercialization, we explore the hypothesis that:

**Hypothesis 1**. Ceteris paribus, female ERC grantees are less likely than male peers to apply for an ERC proof-of-concept grant.

The PoC scheme supports only the early stages of the commercialization process, which is long and complex. A grantee is unlikely to be self-sufficient at the completion of the PoC project, nor to have already reached the market. Therefore, an ERC grantee is unlikely to apply if s/he cannot expect to find support during the following stages of the commercialization process as well. In fact, several studies have revealed that researchers are more likely to engage in commercialization activities such as patenting and licensing when working in an environment with more support infrastructures, such as technology transfer offices (O'Gorman et al., 2008; Perkmann et al., 2013; Sellenthin, 2009; Thune et al., 2016), and at universities with strong interactions with industry (Gerbin and Drnovsek, 2016). In a similar vein, we can expect that grantees will be keener to apply when they work in a country where the level of collaboration in innovation is strong, such as established and frequent research collaborations between public and private bodies. Therefore, we expect that:

**Hypothesis 2**. Ceteris paribus, ERC grantees from countries characterized by more collaboration between public and private bodies in innovation are more likely to apply for an ERC proof-of-concept grant.

Another factor involved is the amount of financial support provided by the PoC, which is expected to be more attractive for some researchers than others. In fact, researchers' strategies for selecting external funds take into consideration the characteristics of the context in which they work (Laudel, 2006). A very important contextual condition that is expected to affect a researcher's likelihood of applying for a PoC grant is the cost of personnel. For example, the PoC's grant of €150,000 would pay for two PhD bursaries in Italy but only half of a PhD bursary in Norway. In other words, the PoC grant is arguably less attractive for researchers working in countries where the cost of personnel is high. Hence, we expect that:

**Hypothesis 3**. Ceteris paribus, ERC grantees from countries with a high cost of personnel are less likely to apply for an ERC proof-of-concept grant.

A second important contextual factor that determines the financial attractiveness of a PoC grant is the existence of alternative funding sources for innovation. European countries differ considerably in their level of public and private investment in innovation, research, and development (European Commission, 2020; Research Council of Norway, 2021). Researchers are expected to evaluate the opportunity cost of applying for a grant by weighing the effort involved in developing a proposal vis-à-vis the chances of winning and the amount of financial support they would receive. They may also consider the opportunity cost of alternative funding sources from private investors and/or national funding programs. As a result, researchers from countries with an abundance of alternative sources may feel less urgency to apply. Hence:

**Hypothesis 4**. Ceteris paribus, ERC grantees from countries with more financial support for innovation are less likely to apply for a PoC grant.

### 2.3. *Factors predicting successful proposals*

Peer review of grant proposals is not exempt of criticism in relation to its reliability, predictive validity, and bias (Abdoul et al., 2012; Bornmann, 2011; Lee et al., 2013; Marsh et al., 2008). Therefore, there is an element of chance in receiving a research grant (Cole et al., 1981; Mayo et al., 2006; Seeber et al., 2021). This section summarizes the results from past research about whether certain characteristics of the consortia, the organization, the individual scientist, or the proposal affect a proposal's chances of success. Next, we develop hypotheses on how three new factors explored in this paper affect the chances of a proposal to win.

#### 2.3.1. *The consortia*

Wanzenböck et al. (2020) explored the factors affecting the success of the consortium of organizations applying to Horizon 2020. They found that proposals from consortia with high levels of experience and reputation, involving a large share of Western European partners, and engaged in more application-oriented research had greater chances of success.

#### 2.3.2. *The organization*

Some studies have explored the characteristics of the institutions participating in the EU's funding programs and found a preponderance of institutions with strong academic reputations and scientific productivity. However, these works did not consider how many proposals they submitted in the first place (Geuna, 1998; Henriques et al., 2009; Lepori et al., 2015).

Other studies explored the predictors of success rates in grant applications. Enger and Castellacci (2016) reported that Norwegian research organizations that had participated previously in Horizon 2020 and had a strong scientific reputation were more likely to succeed again in their Horizon 2020 applications.

Murray et al. (2016) documented that applicants from smaller institutions were consistently less likely to receive grants from the Canadian Natural Sciences and Engineering Research Council (NSERC) Discovery Grant program (2011–2014) and more likely to receive smaller grants. In contrast, Piro et al. (2020) found no impact of institutional size on the success rate. They indicated that institutional predictors of success vary across thematic areas and to some extent over time, reflecting the changing goals of the EU's programs.

#### 2.3.3. *The individual scientist*

Studies have shown that the assessment of grant proposals is influenced by the applicants' past performance, either in terms of scientific production or previous grant awards (Tamblyn et al., 2018). Van den Besselaar and Leydesdorff (2009) found that the final decisions of the Netherlands Research Council for the Economic and Social Sciences correlate with the past performance of applicants with regard to their publications and citations. Studying recent PhD applications to the NWO, Bol et al. (2018) found that early funding success increases the likelihood of obtaining later funding, with winners just above the funding threshold receiving more than twice as much research funding. They also noted that such differences are not due to achievements enabled by the preceding grant. Neufeld et al. (2013) reported a strong overlap in the scientific productivity of funded and non-funded applicants for ERC Starting Grants (Stg), with only a few fields showing a significant difference.

Several scholars explored the presence of gender-related differences. Van den Besselaar and Leydesdorff (2009) discovered that proposals from female scientists are more likely to be funded than might be expected from their past scientific performance and reviewers' evaluations. In contrast, studies of grant funding from the Austrian Science Fund and the NWO found no differences related to the gender of the applicant (Mutz et al., 2012; Albers, 2015; Volker and Steenbeek, 2015).

Sandström and Hällsten (2008) observed that proposals to the Swedish Research Council from principal investigators with the same affiliation as the proposals' reviewers receive a higher grade than applicants with no related affiliations. In addition, among scientists with no related affiliations, female PIs receive a “bonus” of around nine percentage points.

#### 2.3.4. The proposal

Boyack et al. (2018) examined 369 proposals to the National Institutes of Health (NIH) in the US. They reported that a clearly articulated proposal is more likely to be funded. In addition, they established a correlation between the proposal's success and a great deal of topical overlap between the proposal's references and the applicant's prior publications.

Studying the peer review process for medical research grant proposals at a leading medical research university, Boudreau et al. (2012) noted that evaluators systematically give lower scores to research proposals that were closer to their own areas of expertise and to those that are highly novel. In a similar vein, Van den Besselaar et al. (2018) conducted a textual linguistic analysis of the evaluation reports of applications to the ERC StG and found that panels focused on the applications' weak points rather than looking for ground-breaking ideas. Interdisciplinary research proposals consistently had less funding success in the Australian Research Council's Discovery Program (Bromham et al., 2016), while had a similar funding success in the Collaboration in Science and Technology (COST) European program (Seeber et al., 2022).

#### 2.3.5. New variables and hypotheses

It might be argued that proposals that failed once, are intrinsically of lower quality and hence less likely to succeed at successive attempts. However, only some of the unsuccessful proposals are re-submitted, because only grantees that truly believe in the project are likely to continue applying. Therefore, proposals that are re-submitted not only can exploit the feedback from the first round of evaluation, but they will also come from strongly motivated grantees who believe in the proposal's potential. Hence, we formulate the hypothesis that:

**Hypothesis 5**. Ceteris paribus, further attempts of a proposal are more successful than the first attempt.

In the previous section we explained that grantees from countries with high cost of personnel are less likely to apply. It is plausible that grantees from countries with a high cost of personnel are therefore likely to apply only if they perceive to have very high chance of success. Thus, because of such self-selection process, we expect that:

**Hypothesis 6**. Ceteris paribus, proposals of ERC grantees from countries with a high cost of personnel are more likely to be successful.

Finally, we argued that grantees in countries where the financial support for innovation is low are expected to apply more frequently. On the one hand, this may imply that some grantees apply even when they do not have realistic chances of success. On the other hand, we might expect that grantees in countries where the financial support for innovation is low will put more effort in developing a proposal, because they have few or no alternative funding options. This is true especially in countries with very low support for innovation. Therefore, we expect that:

**Hypothesis 7**. Ceteris paribus, proposals of ERC grantees from countries with a very low support for innovation are more likely to be successful.

Table 1 summarizes our current knowledge about the factors predicting a scientist's likelihood of applying for a grant and the research proposals' success, as well as the new variables we examined. Through AML, we will identify the variables that have the strongest impact on the probability of applying for and winning a PoC grant. We will also use AML to determine the relationship between the predictor variables and the targets in the datasets. AML can uncover interactive and non-linear relationships. It increases the probability of finding a model that fits well with the data and that can be generalized across unseen data sets.

## 3. Data and methods

### 3.1. Datasets

This study relies on two distinct datasets: i) the ERC dataset of funded projects and respective principal investigators (PI), and ii) the PoC dataset of applications and granted projects, which were complemented using specific sources.

#### 3.1.1. The ERC dataset of funded projects

The European Research Council (ERC) is a public body established by the European Commission in 2007 to encourage the highest quality research in Europe through competitive funding and to support bottom-up, investigator-driven frontier research across all fields, with proposals being selected solely based on their scientific excellence. The ERC's funding schemes have been part of the EU's Seventh Research Framework Program (FP7) (2002−2013) and Horizon 2020 (2014–2020) programs dedicated to research and innovation.

The dataset includes information on all of the 10,074 projects the ERC funded through the Seventh Framework Program (FP7) and Horizon 2020 (H2020) with a starting date prior to January 2020. The dataset includes the *year of the call for proposals*, *type of grant*, *project*: *title*, *abstract* and *acronym*, *scientific panel*, *project's start date*, *end date* and *duration*, *grantee's institution of affiliation* and *country of affiliation*, *grantee's first name and surname*, and *gender* (see Section 3.3 for more details).

#### 3.1.2. PoC proposals

The ERC's Proof of Concept funding was introduced into the ERC Work Program in 2011 with the objective of supporting ERC grantees (principal investigators - PI) in establishing a proof-of-concept or identifying a development path or intellectual property rights strategy for ideas arising from their past or current ERC funded project. The goal of the funding is to help the grantees bring ideas originating from an ERC grant to a pre-demonstration stage, and look for investors (venture capitalists, companies, or social entrepreneurs) interested in commercialization or roll- out of the technology or product, depending on the nature of the invention or idea. It is possible to be awarded more than one PoC grant per ERC funded project, but only sequentially, not cumulatively, meaning that the same ERC funded project can only have one active PoC grant at a time. Eligible PIs are those with an ongoing ERC grant or one that has ended not >12 months prior to the PoC call for proposals' publication date or 1 January of the call year. Each grant is usually close to 150,000€ for a 12–18-month period and cover activities at the very early stage of turning research outputs into a commercially or socially valuable proposition, meaning the initial steps of pre-competitive development.[1]

The dataset includes 2186 proposals from 2011 to 2017, with information on the *call for proposal year*, *host institution* and *country*, *project*

[1] Funding can be used to establish viability, technical issues, and overall direction, clarify intellectual property rights, position and strategy, provide feedback for budgeting and other forms, provide connections to the later stage funding of commercial projects, and cover initial expenses for establishing a company.

**Table 1**
Predictors of a scientist's likelihood of applying for a grant and the research proposal's success: Existing evidence and new variables.

A scientist's likelihood of applying for a grant

| Level | Factor | Result | Context |
|---|---|---|---|
| Organization | Complementary national funding schemes | Positive | Norway |
| Organization | Prior participation | Positive | Norway |
| Organization | Strong network in collaborative research | Positive | EU |
| Individual | Prior success | Positive | Netherlands |
| Individual | Past scientific performance | Positive | EU |

A scientist's likelihood of applying for a grant

| Level | New explored factor | Expected effect |
|---|---|---|
| Individual | Gender | Positive (male) |
| Country | Cost of personnel (high average wage) | Negative |
| Country | Financial support for innovation (*Finance and Support*) | Negative |
| Country | Collaboration in innovation (*Linkages*) | Positive |
| Individual | Research panel | (Exploratory) |
| Institution | Institution of affiliation | (Exploratory) |

A research proposal's likelihood of research proposal success

| Level | Factor | Result | Context |
|---|---|---|---|
| Consortium | Experience, reputation, West European, applied orientation | Positive | EU |
| Organization | Size | Positive, no effect | Canada, EU |
| Organization | Prior participation | Positive | Norway |
| Organization | Reputation | Positive | Norway |
| Individual | PI's publications overlap with proposal's content | Positive | US |
| Individual | Gender | Positive (female) or no effect | Netherlands, Sweden, Switzerland |
| Individual | Past scientific performance | Positive, no effect | Netherlands, EU |
| Individual | Early grant success | Positive | Netherlands |
| Individual | Same affiliation as the reviewer | Positive | Sweden |
| Proposal | Quality of writing-clarity | Positive | US |
| Proposal | Novelty | Negative | US |
| Proposal | Interdisciplinarity | Negative | Australia |
| Proposal | Weak points more important than ground-breaking ideas | Negative | EU |

A research proposal's likelihood of research proposal success

| Level | New explored factor | Expected effect |
|---|---|---|
| Proposal | Attempt number | Positive |
| Individual | Research panel | Partly related to propensity to apply (Exploratory) |
| Organization | Institution of affiliation | (Exploratory) |
| Country | Cost of personnel (high average wage) | Positive |
| Country | Very low financial support for innovation (*Finance and Support*) | Positive |

*title* and *project abstract*.

### *3.2. Matching datasets*

The goal of the analysis is to predict i) which ERC grantees applied at least once for a PoC grant, and ii) which proposals won.

The two datasets include complementary information. To combine them, we matched ERC grantees and projects to a specific PoC proposal. Most of the matches were unproblematic, because the ERC dataset included winning PoC proposals as well as several ERC projects (StG, CoG and AdG) with the same title and acronym as the PoC proposals, or the abstract of the PoC proposal included an ERC's project title, acronym or code of the ERC grant or the name of the grantee. Next, we examined pairs of PoC proposals and ERC projects[2] written by scientists affiliated with the same institution. We matched them when: i) they were the only proposals from a given institution, ii) they had the highest text similarity score based on explicit semantic analysis (Gabrilovich and Markovitch, 2007) and the grantee of the first, second and/or third most similar ERC was the same; and iii) the text similarity was the closest by a wide margin (>0.100 on a scale of 0 to 1).

Applying these rules, we identified: i) 1166 of the 6697 ERC grantees between 2007 and 2016, who had applied at least once for a PoC grant, and ii) the authors of 1878 of the 2186 PoC proposals (86 %) – with a similar level of coverage across countries, such as, for the three most represented countries: 84 % for United Kingdom, 87 % for the Netherlands and 86 % for Germany. It is important to highlight that in principle an ERC grantee's affiliation can differ from its' PoC affiliation. In practice, this happened in a very small number of cases, in fact only 6 % of PoC grantees were based at a different Host Institution than the ERC grant and only 1 % in a different country. Due to the anonymity of the data, we cannot fully correct these cases. Nevertheless, given such a small number, we can expect a negligible effect on the results.

[2] We also took into consideration that the year of the PoC had to be more recent than the year of the ERC call for proposals, and the year of the PoC call for proposals could be no later than 1 year after the end of the ERC grant.

### 3.3. Variables

#### 3.3.1. Dependent variables

*3.3.1.1. ERC grantee applying for a PoC grant.* This is a binary variable that takes value "1" if an ERC grantee has applied at least once for a PoC grant and "0" is not. Of the 6697 potential applicants, there were 1166 actual applicants (17.41 %).

*3.3.1.2. Winning proposal for PoC grant.* This is a binary variable that takes the value "1" if the proposal was successful (665 of 1830 proposals, or 36.34 %) and "0" if not successful.

#### 3.3.2. Predictive variables

Most of the predictive variables are available through the dataset of ERC grantees and are used as predictors of both the likelihood of applying for a PoC grant and of a PoC proposal being successful.

*3.3.2.1. Scientific panel.* In the Seventh Framework Program and Horizon 2020 there were 25 scientific disciplinary panels that evaluated the research proposals: nine in the life sciences, ten in the physical sciences and engineering, and six in the social sciences and humanities.

*3.3.2.2. Country.* The country in which the ERC grantee works (host country).

*3.3.2.3. Cost of personnel.* We expect that part of the variations between countries in the likelihood to apply is due to differences in the cost of hiring research personnel (e.g., PhD students, postdocs). As a proxy for the cost of personnel we used the nominal average annual wage for 2010–2017 in US$ (source: OECD - https://data.oecd.org/earnwage/average-wages.htm). This information is missing for the 33 grantees from Serbia, Cyprus, Croatia, Romania, and Bulgaria.

*3.3.2.4. Type of ERC grant.* There are three different types of ERC grants: i) Starting grants – StG (4307 grants), designed to support outstanding researchers of any nationality who are starting to develop an independent career (2 to 7 years of experience since completion of their PhD), ii) Consolidator grants - CoG (1905 grants), established in 2013, designed to support researchers at the stage of consolidating their own research team or program (7–12 years of experience since completion of their PhD), and iii) Advanced grants – AdG (2855), which support well-established and outstanding scientists pursuing innovative, ground-breaking research.[3] For multiple grants winners, we considered the first grant won as a predictor of the likelihood of applying for and winning another grant.

*3.3.2.5. Year of the ERC call for proposals.* The dataset includes calls from 2007 to 2019. The CoGs were introduced in 2013, whereas the PoC grants were introduced in 2011. This is an important control, because later winners might be in the early stages of their project and hence less likely to apply for a PoC grant.

*3.3.2.6. PI's gender.* The gender of those applying for the grant: male or female.

As a proxy for the *financial support for innovation* available in each country, we considered the composite indicator "*Finance and support*" from the European Innovation Scoreboard (EIS),[4] which measures the availability of finance for innovation projects using venture capital expenditures, and the support of governments for research and innovation activities using R&D expenditures in universities and government research organizations. We used the average score between 2012 and 2017.

The "*Linkages*" composite indicator from the EIS measures the country's level of *collaboration in innovation*. It considers the collaboration efforts between innovating firms, research collaboration between the private and public sector, and the extent to which the private sector co-finances public R&D activities. We used the average score between 2012 and 2017.

*3.3.2.7. Institution of affiliation.* The organizational environment in which a scientist works may affect his/her likelihood of applying for and winning a grant. In the last quarter century, several scholars have highlighted the increasing efforts of university leaders to nudge researchers toward an entrepreneurial orientation (Clark, 1998). The efficacy of such efforts has varied across countries and systems (Seeber, 2013), perhaps implying that the organizational level is not always an important one.

### 3.4. The ERC's PoC evaluation process

A PoC proposal includes: i) a short description of the idea used to develop a proof of concept, and the link between the initial ERC funded project and the proposed idea; ii) an outline of the innovation strategy or potential of the idea and how it will lead to a commercial or social innovation; iii) an outline of the economic and/or societal impact expected from the project; and iv) an outline of a reasonable and plausible plan of activities supporting the feasibility of the project.

The evaluation process consists of a single-stage submission and single-step evaluation involving five reviewers per proposal (independent experts), who know the identity of the applicant (i.e., a single blind review). Reviewers evaluate each eligible proposal independently and grade it as "pass" ("very good" or "good") or "fail" for each of the three evaluation criteria: i) excellence of the innovation potential; ii) impact of the project; and iii) quality and efficiency of the implementation.

To be considered for funding, a proposal must be awarded a passing grade ("very good" or "good") by most reviewers on each of the three evaluation criteria. A proposal that fails to meet one or more of the criteria will not be funded. If there is not enough money to fund all of the proposals that pass all three evaluation criteria, the proposals that do pass will be ranked and funded in order of this ranking. Unlike the other ERC grants (StG, CoG, AdG), the proposals are *not* grouped into separate disciplinary panels or domains. If necessary, the reviewers will meet as an evaluation panel to determine a priority order for proposals that have the same ranking, although, to date, this situation has not occurred.

### 3.5. Methods

As mentioned in the introduction, we used AML to explore the structure in the data and the predictors. Indeed, our innovative approach is one of our empirical contributions. Specifically, we used DataRobot as the software program, following the guidelines written by Larsen and Becker (2021). DataRobot is one of the leading software developers in the field of AML with similar capabilities to those best in class (for a review of different software solutions, see Truong et al., 2019). DataRobot runs on AWS, GCP, and Azure platforms and can process numerical, categorical, and time-series data.

The lack of a shared understanding about the exact predictors and the shape of the relationships between these predictors and the target variables (applying for and winning a grant) provides fertile ground for using AML, because it can deal with big data and complex relationships in ways that minimize predictive errors (Kuhn and Johnson, 2013).

The purpose of AML is to explore patterns in the data and to make predictions based on these patterns in an empirically robust way. The

[3] The dataset of ERC grantees also includes information on 1007 PoC grants.

[4] The EIS is a dataset providing a comparative assessment of the research and innovation performance of the EU member states and the relative strengths and weaknesses of their research and innovation systems.

common AML pipeline includes: 1) data pre-processing and feature engineering, 2) model selection, hyperparameter optimization, and architecture search, and 3) model interpretation and predication analysis (Truong et al., 2019).

Data pre-processing and feature engineering is the first step in the AML process, after choosing the target variable or dependent variable. It requires considerable researcher intervention with respect to the identification of the dependent variable, potential leakages, and selection of the relevant predictors. Data leakages in machine learning happen when algorithm training is done with data otherwise not available at the point of prediction or data one tries to predict as part of the predictors. The selection of potential variables should be done on the basis of available theory and practice (as we explained in the literature review section). At this stage, redundant variables may be deleted, and informative features derived. Informative features are the independent variables that can uniquely predict the target or dependent variable.

Data pre-processing is one of the innovations of machine learning. It uses encoders to convert data for the best use in different models, allowing the machine to detect basic data types or schemas, and to help with feature engineering from user input specifications (Truong et al., 2019). Examples include converting categorical data into integers, imputing missing data, and standardizing the data. Different pre-processing modules can be used for different types of models. DataRobot attempts different data transformations, missing variable extrapolations, and pre-processing modules when preparing for model selection and application. Different models in the next step might also require different pre-processing procedures. For example, some models may not work well with binary data or various distributions. Therefore, in the pre-processing stage different algorithms are applied to treat the data before modeling.

The second step is model selection. AML's model inventory includes logistic regressions, tree-based algorithms, SVM, neural network models, genetic algorithms, Bayesian searches, and random searches. Many different model types are tried with many different sets of parameters or hyperparameters to find the best model.

Data are partitioned into several parts, and the AML attempts to find the best model for each data part, which is then tested on unseen parts of the data. DataRobot uses a stratified partitioning method with 5-fold cross validation and a 20 % holdout sample. DataRobot considers 16 % of the sample and uses that information to determine which models to run with 32 % of the sample (Larsen and Becker, 2021). This process of iteration repeats itself with different models. Some models combine multiple best-in-class ("blender") models to achieve accurate predictions. The best models are the ones that ultimately minimize the prediction errors across the different data subsets, as well as the holdout set. Machine learning models utilizing cross-validation and that have the freedom to explore complex interactive and non-linear relationships can minimize the errors associated with the bias variance dilemma using a minimum number of variables that explain the most variance across different parts of the data (Doornenbal et al., 2021). DataRobot works with informative features (variables that contribute to predictions in unique ways) to construct multiple models. Given that our data have binary outcomes, we used logloss accuracy metrics to assess predictive accuracy. We designated 20 % of the data as the holdout sample and divided the rest of the data into five folds. The result of the combined total of each of the five validation partitions is the cross validation.

The third and final step is model interpretation and predication analysis, which is the key outcome of AML. It provides a detailed representation of the results through a model dashboard comparing different models and recommending the most accurate one, identifies the importance of various features, and presents various visualizations that we will illustrate in the results section. Following the recommendations from Larsen and Becker (2021), in our case we chose the models with the best logloss metrics in the validation, cross-validation, and holdout samples.

## 4. Results

### 4.1. Descriptive analysis

Fig. 1 shows that the number of new ERC grantees[5] has grown from 294 in 2007 to over 800 after 2012. The number of ERC grantees applying for a PoC also increased until 2012 and then declined. The reason is that most projects starting after 2012 were still ongoing in 2017, the last year we considered for PoC applications.

There are large variations between the grantees of different scientific panels in their likelihood of applying for and winning a PoC grant – even between panels within the same research domain. The likelihood of applying for and winning a PoC grant are strongly correlated (0.65). Thus, panels with more/less likelihood of having grantees who apply for grants are also those that have proposals that are more/less likely to win compared to the average. For example, grantees in *SH6 The Study of the Human Past* are −45 % less likely to apply, and proposals in this panel are 92 % less likely to win. Similar results occur for *SH1 Individuals, Markets and Organizations* (−85 %, −69 %), *PE1 Mathematics* (−68 %, −45 %), and *PE9 Universe Sciences* (−68 %, −31 %). Other fields display the opposite trend. They have candidates who are more likely to apply and proposals that are more likely to win. Examples include *PE5 Synthetic Chemistry and Materials* (+58 %, +47 %), *PE7 Systems and Communication Engineering* (+100 %, +26 %), and *LS7 Applied Medical Technologies, Diagnostics, Therapies and Public Health* (+70 %, +23 %). As a result, there is a large difference between panels in their share of ERC and PoC grantees. For example, *PE4 Physical and Analytical Chemical Sciences* represents 4.2 % of ERC grantees and 14.6 % of PoC grantees, and *LS7 Applied Medical Technologies, Diagnostics, Therapies and Public Health* represents 5.2 % of ERC grantees and 11.1 % of PoC grantees, whereas all panels from the *Social Science and Humanities* domain represent 20 % of ERC grants but just 6.9 % of PoC grants.

Remarkable differences in the likelihood of applying exist across countries as well. Fig. 2 illustrates the top 10 European countries by number of ERC grantees and identifies those who applied at least once. Grantees from Spain (25 %), Israel (24 %), Italy (22 %), and the Netherlands (21 %) are more likely to apply, whereas grantees from Switzerland (13 %) and Germany (14 %), France, and Sweden (15 %) are the least likely.

Differences by country in the likelihood to apply an in the chances of winning lead to a different geographical distribution of PoC grants compared to ERC grants. For example, the UK, Germany, France, and Switzerland together account for 56 % of the ERC grants in 2007–2016, and 45 % of the PoC grants in 2010–2017. In contrast, Mediterranean countries such as Italy, Israel, Greece, and Spain together won 17 % of the ERC grants and 28 % of the PoC grants. At first glance, this pattern may be surprising, given the strong tradition in applied research and entrepreneurship of the first group. However, precisely because of such a tradition, scientists in such systems may have funding opportunities for translational research from other public and private sources, whereas the PoC program addresses the stronger demand for funding support in the second group of countries that have fewer such opportunities.

Winners of CoG apply less frequently (4 %) compared to StG and AdG (17 % and 21 %) – the reason being that CoG really began only in 2014 and most PIs apply in the later stages or after the end of their project. However, the success rates of winning the grant are similar: 37 % for StG, 43 % for CoG, and 34 % for AdG. Men are much more likely than women to apply (17.4 % vs. 12.7 %) as well as to win (37.2 % vs. 31.8 %). Variations across institutions of affiliation are minor (see the next section).

To disentangle the predictive power (importance) of each variable, we conducted an inferential analysis.

[5] We counted a grantee only once, the first time s/he received a grant.

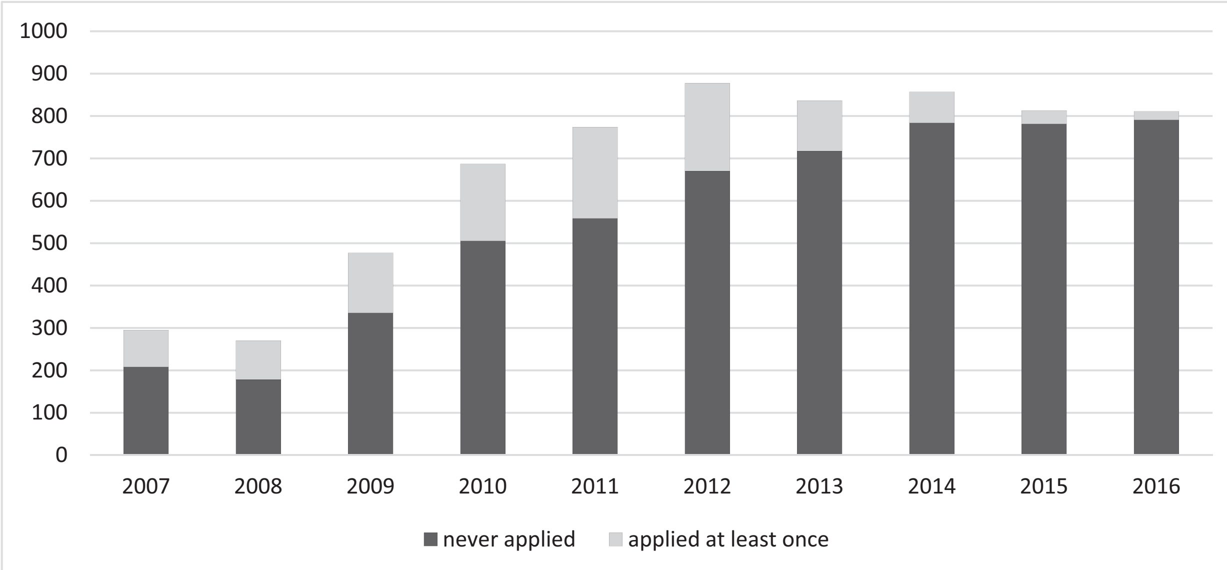



**Fig. 1.** ERC grantees choosing to apply for a PoC grant (by year of first grant, 2007–2016).

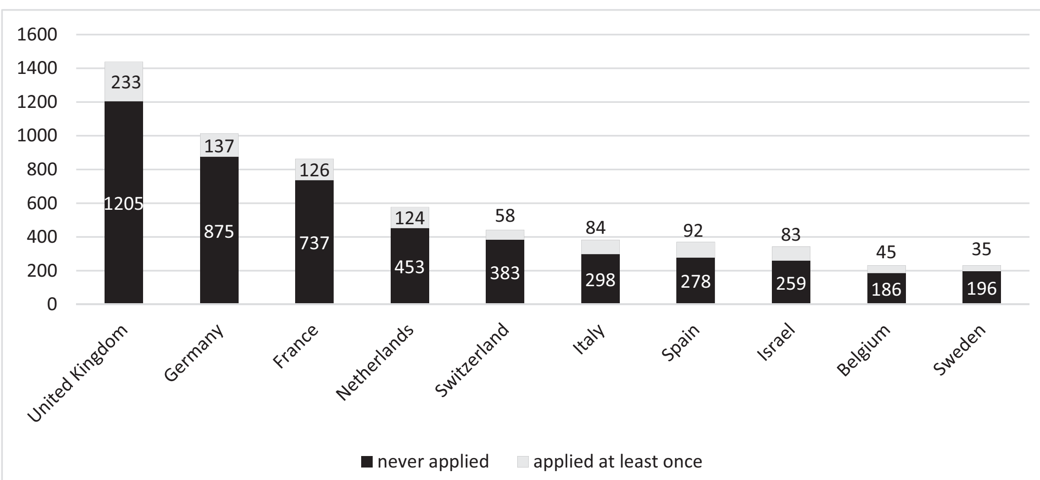



**Fig. 2.** Likelihood of applying for a PoC grant: Top 10 countries by number of ERC grantees.

### 4.2. Inferential analysis

#### 4.2.1. Likelihood of applying

Which traits of the country and the applicant predict the likelihood of the 6697 ERC grantees applying at least once for a PoC grant between 2007 and 2016? To answer this question, we ran an AML analysis of our first dataset and generated the blueprint in Fig. 3. Altogether, 97 potential models were evaluated, each of which uses a different set of pre-processing steps unique to that model. We chose the ENET blender model due to its relative performance and the fact that we were trying to evaluate all of the relevant features across different models. The ENET blender model selected has a logloss of 0.389, 0.389, and 0.381 for the validation, cross-validation, and holdout samples, respectively.

The ENET model is a blender model, meaning, a combination of several models simultaneously. In this first case, the ENET blended three different Nystoem Kernel SVM classifier models, each with its own data pre-processing. This class of models approximates the Kernel Support Vector Classifier using a sklearn Logistic Regression, a machine learning classification algorithm that is utilized to predict categorical dependent variables. Support Vector Machines (SVM) are a class of "maximum margin" classifiers that seek to maximize the separation they find between classes. Optionally, they can include a penalty function that allows them to mis-classify some observations for the sake of wider margins between the classes for the rest of the observations.

SVMs are very efficient in high-dimensional spaces. They make use of a "kernel" function, which allows for a non-linear transformation of the data before fitting SVM for better modeling accuracy. In machine learning, a kernel is a method to apply linear classifiers to non-linear data by mapping data to higher dimensional space. The number of features in a dataset represents the dimensionality. AML uses higher dimensional spaces to model datasets with many attributes, with each record or observation represented as a point in space with its position on the attribute values. A feature space can be a collection of features related to some properties of the object. Kernel map approximations allow SVMs to run much faster and scale up to larger datasets (Chang and Lin, 2011).

We then analyzed the predictions further using an elastic-net classifier with binomial deviance to obtain the final prediction. ElasticNet is an extension of logistic regressions where the optimizer tries to find a parsimonious model by preferring simpler models. A simpler model is

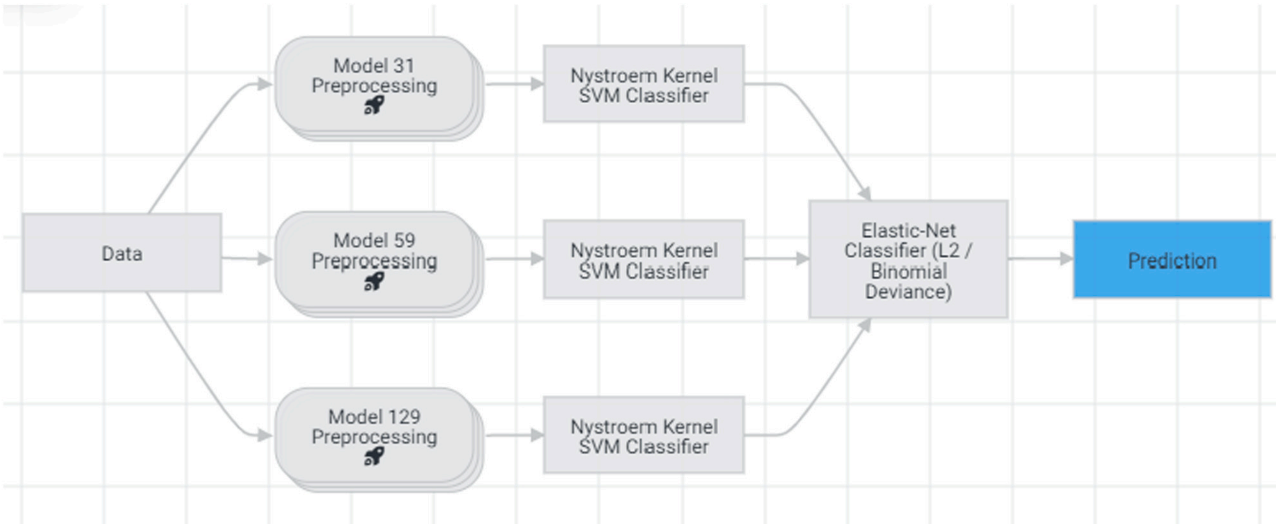



**Fig. 3.** Blueprint for ENET blender model, target applied for a grant.

defined as having coefficients with smaller absolute values as well as fewer non-zero coefficients. These attributes can help the model deal with co-linear variables and can also produce models that are less prone to overfitting and generalize better to new data. This "preference for simpler models" is formally defined as "regularization." The degree of regularization for non-zero coefficients and the absolute value of the size of the coefficients are the two major meta parameters that control the model. Elastic-net is useful when there are multiple features that are correlated with one another. While Lasso[6] (least absolute shrinkage and selection operator) is likely to pick one of these at random, elastic-net is likely to pick both. A practical advantage of trading-off between Lasso and Ridge[7] is that it allows Elastic-Net to inherit some of Ridge's stability under rotation (Blondel et al., 2013).

Fig. 4 shows the variables' importance, which can be interpreted through the process of perturbation (Fisher et al., 2019; Doornenbal et al., 2021). This process involves adding random noise to the original values of the predictor variable and observing how it changes the prediction and residual errors. Large changes to residual noise show relevance to the predicted value. We used an iterative perturbation process for each of the predictors and tested the "loss drop" in the test set, meaning, an increase in the quadratic mean of the difference between the predicted likelihood (from 0 to 100) and the actual one.

Fig. 5a–h illustrate the exact shape of the relationship between each important variable and the dependent variable in order of predictive power.

Overall, the year of the ERC call for proposals is the most important predictive variable. This result is not surprising, given that later winners are in the early stages of their project and are much less likely to apply for a PoC (Fig. 5a). Disciplinary differences are very important and accord with the observations in the descriptive analysis (Fig. 5b). In line with Hypothesis 1, female ERC grantees are (slightly) less likely than their male peers to apply for a PoC grant (Fig. 5g). This is an interesting result, because a female scientist who obtained an ERC grant could have more motivation and face less constraint to apply for funding in respect to a female scientist who search for funding starting from zero. The fact that, nonetheless, male ERC grantees have a stronger propensity to apply, suggests that - while winning an ERC grant reduces the salience of some of the factors that cause female scientists to participate less in commercial and technology transfer activities, such as opportunity structures and structural barriers, it does not reduce the salience of other factors, such as different attitudes toward commercialization.

In line with Hypothesis 2, working in countries with a higher level of collaboration in innovation increases an ERC grantee's chances of applying. However, there are diminishing returns to scale, suggesting that grantees are discouraged from applying especially when the level of collaboration in innovation is very low (Fig. 5d). Interestingly, there is an inverted U-shaped relationship between the cost of personnel (average wages) (Fig. 5e) and applying for a grant. Only above a certain threshold do higher wages reduce the chances of applying (supporting Hypothesis 3), arguably because over such a threshold it may become difficult to hire research personnel for a sufficient period. For example, a PoC grant is still sufficient to pay for a three-year PhD student in a medium wage country like Italy, but not in a high wage country like Norway. In line with Hypothesis 4, there is a negative relationship between a country's financial investment in innovation and the likelihood of applying for the grant (Fig. 5c). After controlling for these country-level traits, grantees from the Netherlands, Finland, and Greece are substantially more likely to apply, whereas those from Norway are less likely to do so (Fig. 5c). As Fig. 9 indicates, there are also significant variations between institutions, along with the likelihood of winning the grant.

### 4.2.2. *Likelihood of winning a grant*

We followed a similar process for winning grants. We examined 90 possible models. We chose the ENET blender due to its predictive capacity, with a logloss of 0.63, 0.62, and 0.64 for the validation, cross validation, and holdout results, respectively. The blended model provides a meta result of multiple best performing models, highlighting the most relevant and important variables. Fig. 6 shows the blueprint for the model. The ENET model of winning proposals is different from that for applying for grants, as it is based on three different gradient boosted tree classifiers.

Gradient boosting machines (or generalized boosted models) are an advanced algorithm for fitting extremely accurate predictive models and require very little data pre-processing. The model has been called the "Swiss army knife" of predictive models (Friedman, 2001). Gradient boosting machines are a generalization of Freund and Schapire's AdaBoost algorithm (1995) that handles arbitrary loss functions. They are very similar in concept to random forests, in that they fit individual

[6] Lasso is a regression analysis method that performs both variable selection and regularization to achieve better prediction and interpretation.

[7] Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.

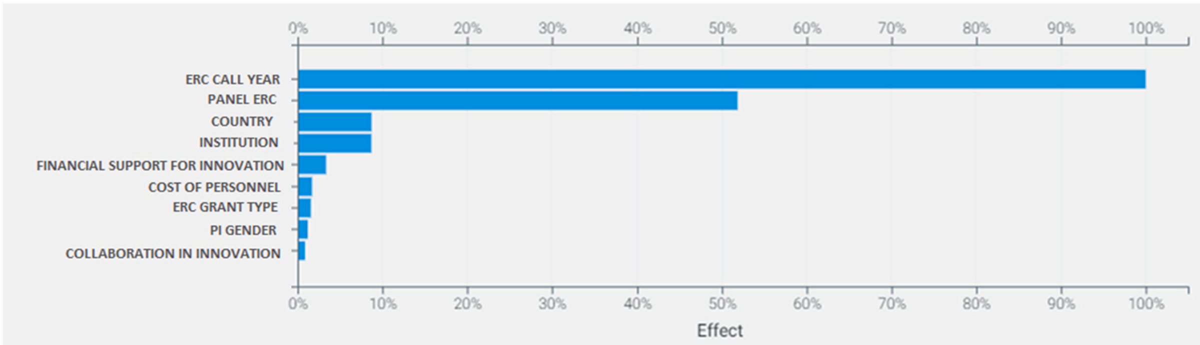



**Fig. 4.** Feature impact, likelihood of applying for a grant.

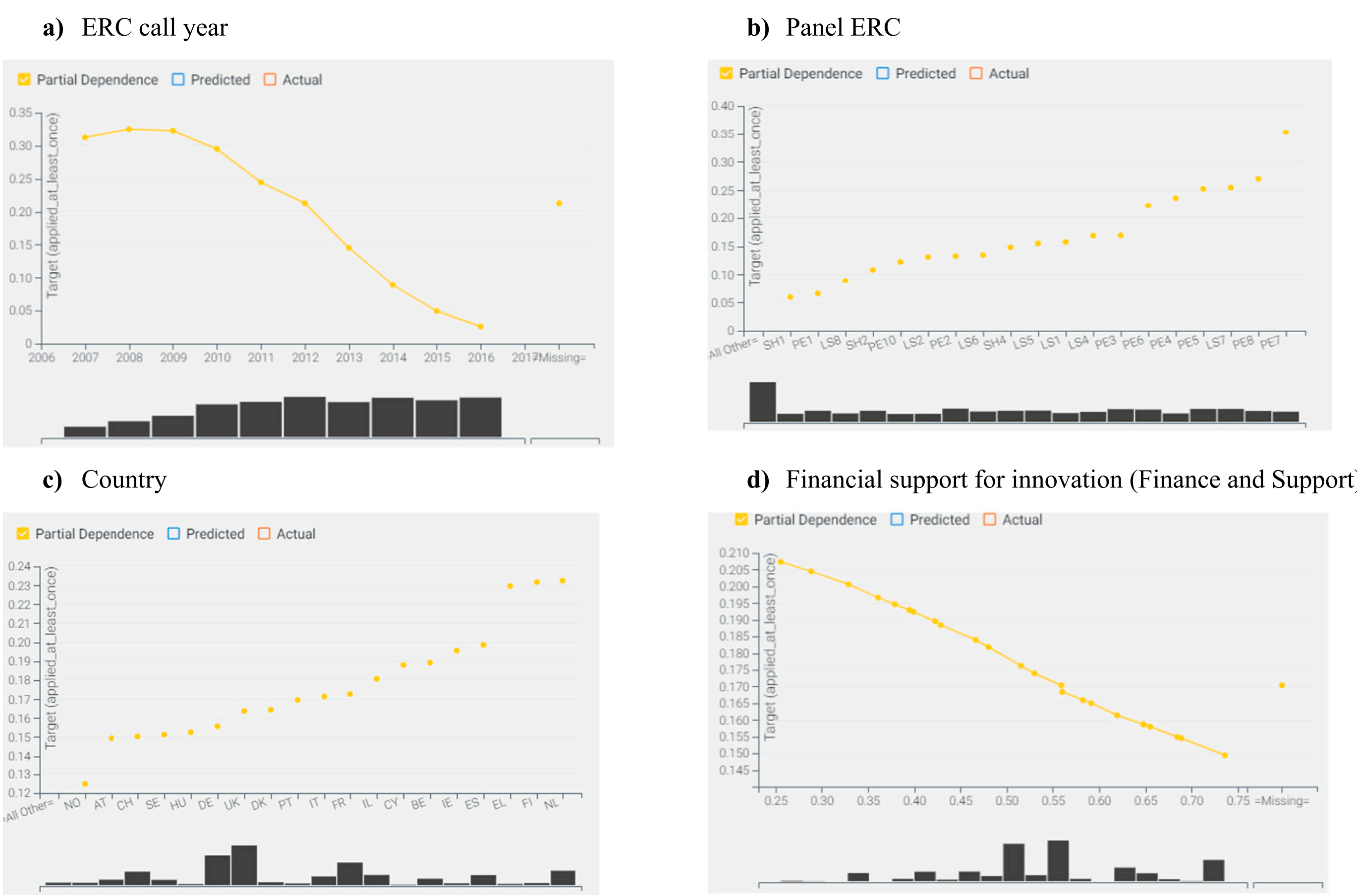



**Fig. 5.** a–d: Relationships between important features and applying for a grant. e–h: Relationships between important features and applying for a grant.

decision trees to random re-samples of input data. Each tree sees a bootstrap sample of the rows of the dataset and N arbitrarily chosen columns, where N is a configurable parameter of the model. Gradient boosting machines differ from random forests in a single major aspect: rather than fitting the trees independently, the gradient boosting machine fits each successive tree to the residual errors from all of the previous trees combined. This difference is advantageous, as the model focuses each iteration on the examples that are most difficult to predict (and therefore the most useful to get correct). Extreme gradient boosting (XGBoost) is a very efficient, parallel version of gradient boosting machines that is very similar to them in R or in Python, but it has been optimized and tweaked by DataRobot for faster runtimes and more predictive accuracy.

Following the same perturbation process described above, we determined the feature importance shown in Fig. 7.

The winning model also shows non-linearity on selected variables (Fig. 8a–g). Disciplinary differences are quite remarkable and confirm evidence of the descriptive analysis. The grantees' disciplines that are more likely to apply are also more likely to win (Fig. 8a). A possible explanation is that research in fields with a stronger potential for commercialization – and hence more likely to apply – often requires extensive resources such as laboratories and/or have easier access to private funding sources. Hence, ceteris paribus, they are less likely to apply and are funded more frequently. In line with the Hypothesis 5, we

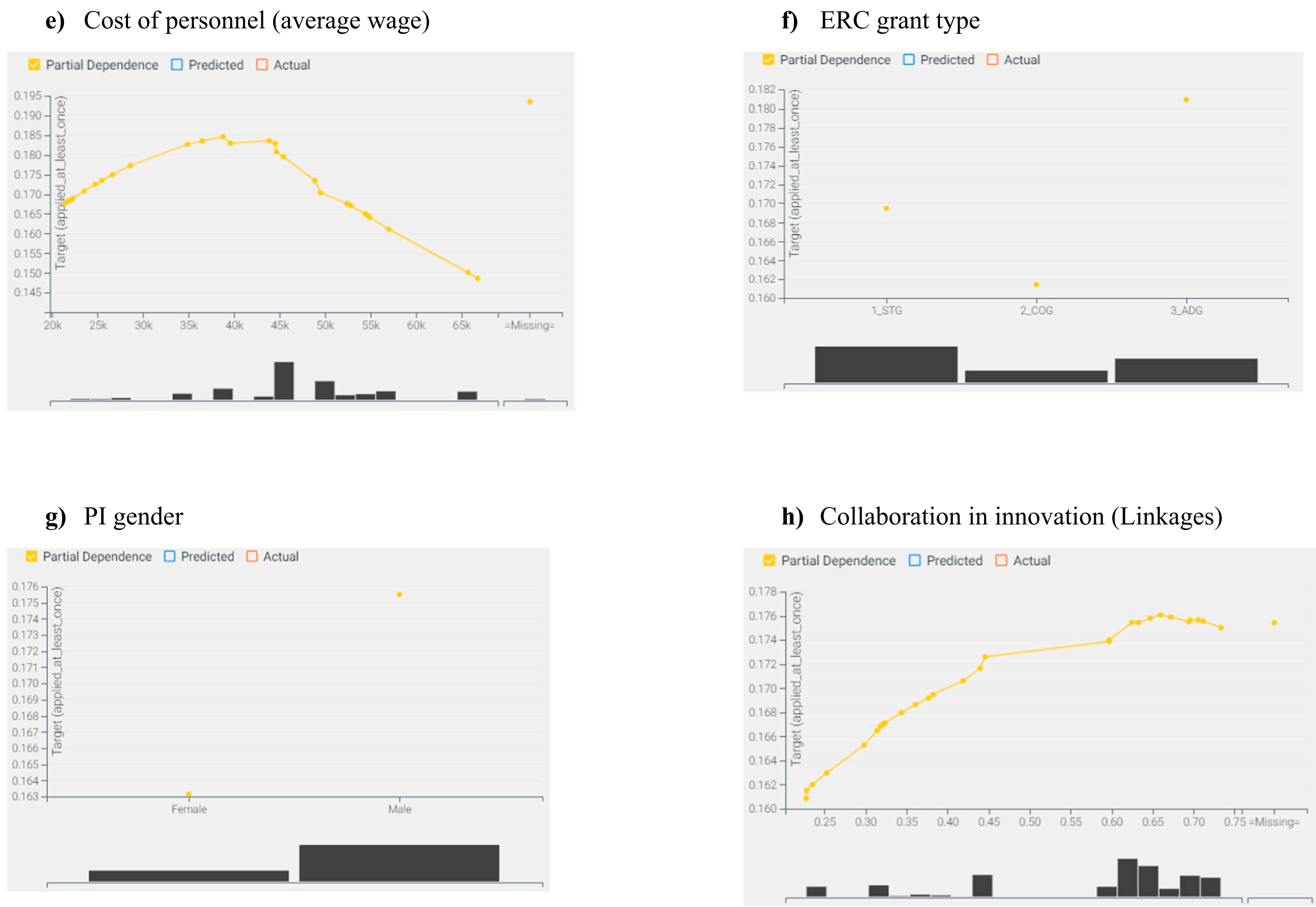



**Fig. 5.** (*continued*).

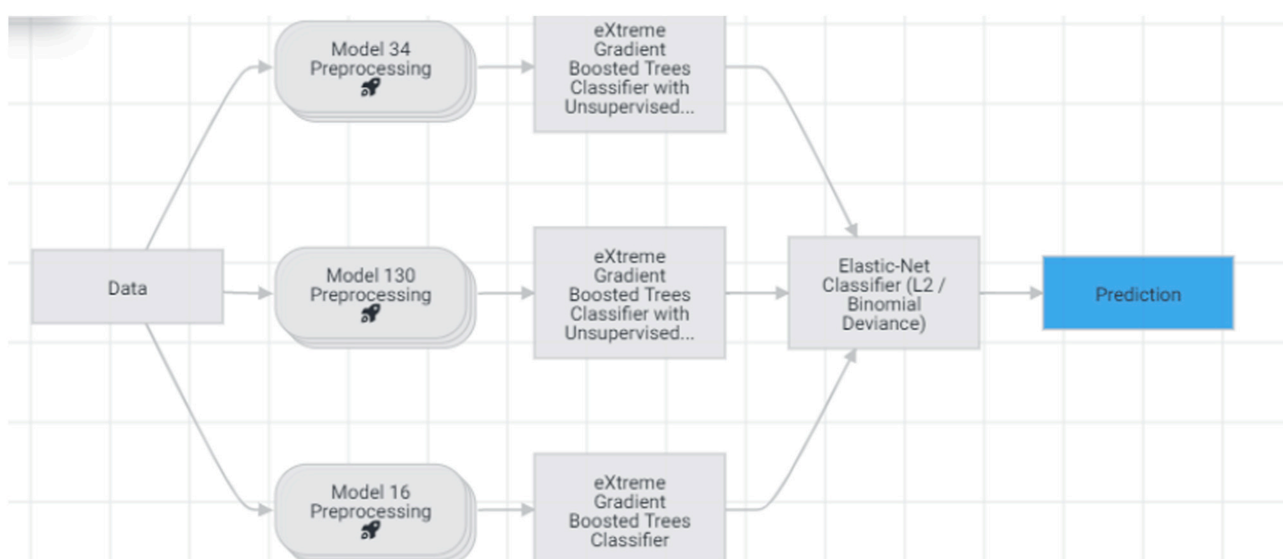



**Fig. 6.** Blueprint for ENET blender model, target winning a grant.

find that second, third and fourth attempts are more successful that the first attempt (Fig. 8c). Advanced grantees are less likely to win (Fig. 8c–d). Ceteris paribus, some countries' proposals are more frequently successful (Fig. 8e). Proposals from high-wage countries are more successful, supporting Hypothesis 6 (Fig. 8f). Finally, proposals from countries with very little investment in innovation are more likely to succeed, backing Hypotheses 7 and the idea that the lack of alternative funding sources provides applicants with a strong incentive to put in the effort needed to create a competitive proposal (Fig. 8g).

Fig. 9 illustrates variations between the largest institutions by the

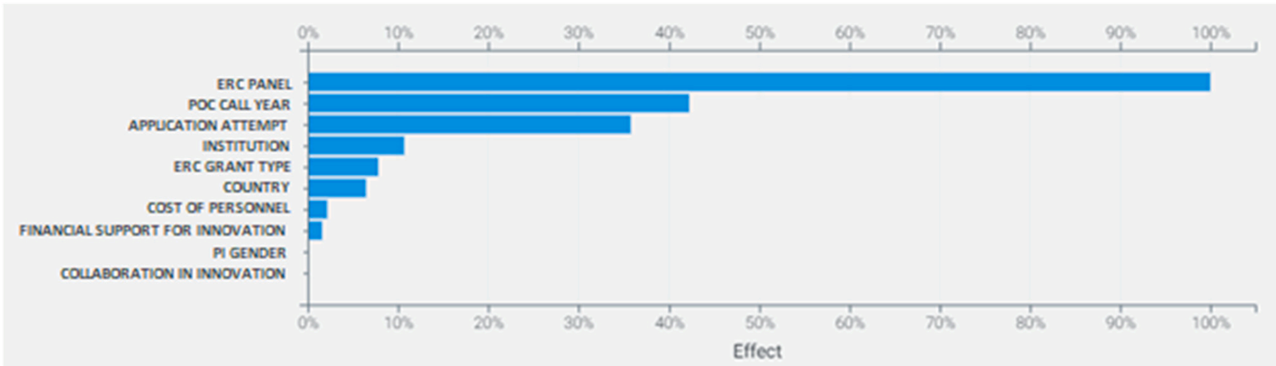



**Fig. 7.** Feature impact, likelihood of winning a grant.


**Fig. 8.** a–d: Relationships between important features and winning a grant. e–g: Relationships between important features and winning a grant.

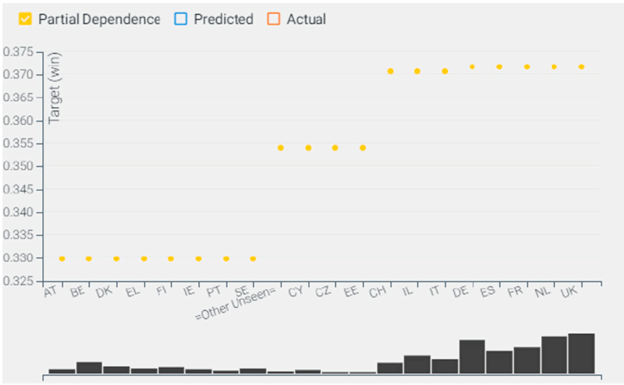
number of applying scientists in the likelihood of applying for and winning a grant. There are some important variations, particularly regarding the likelihood of applying, even between apparently similar institutions such as Oxford and Cambridge Universities, and the Polytechnics of Lausanne (EPFL) and Zurich (ETHZ).

## 5. Discussion and conclusion

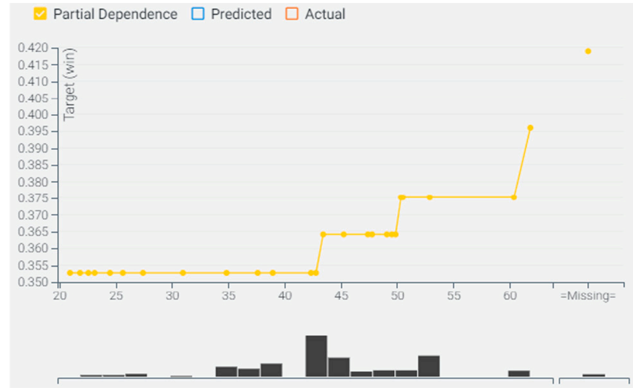
Governments and public agencies have recently introduced funding schemes to support the translation of research activity into commercial applications and address the funding gap that often prevents this transition from happening (Bradley et al., 2013; Munari et al., 2018; Rasmussen, 2008). The Proof of Concept (PoC) funding scheme was established in 2011 by the European Research Council (ERC) to promote the early stages of the valorization process of the research conducted by ERC grantees.

While several studies have explored the variables that predict the success of a research proposal, there is almost no research regarding the factors that affect who applies for a grant. Therefore, we explored the factors that predict which ERC grantees applied for a PoC grant and which PoC proposals were successful by combining information from two datasets of 10,074 ERC grants (representing 8361 individual

**e)** Country




**f)** Cost of personnel (average wage)




**g)** Financial support for innovation (Finance and Support)

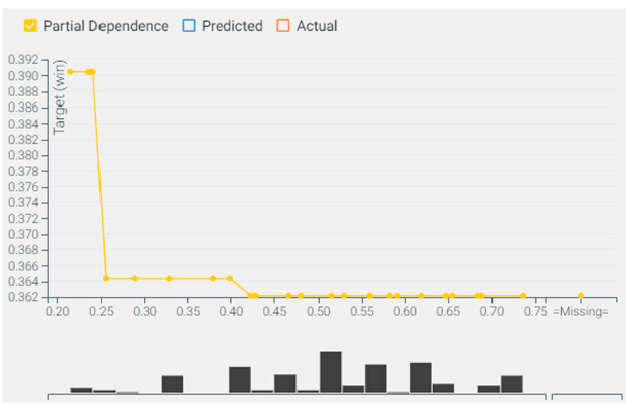



**Fig. 8.** (*continued*).

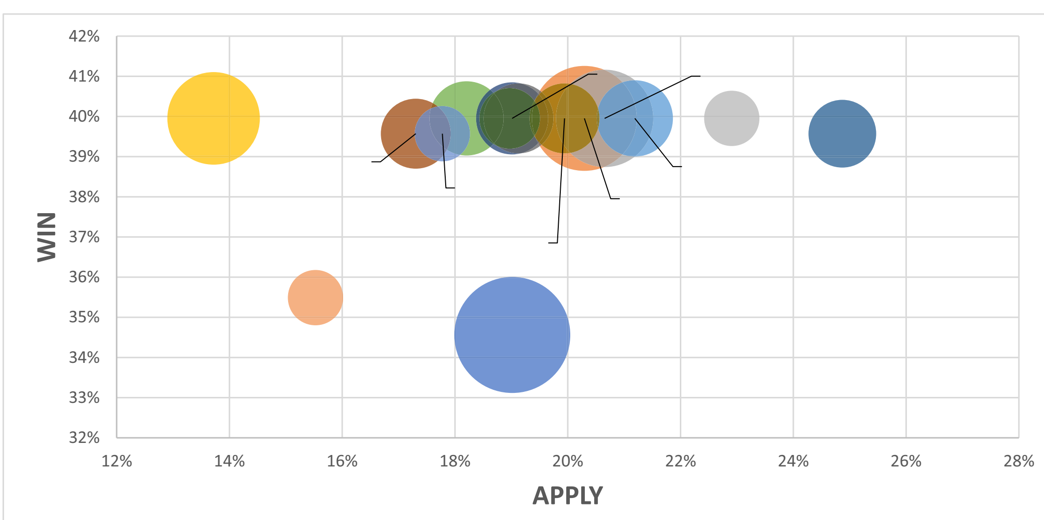



**Fig. 9.** Institutional likelihood of applying for and winning a grant.
Legend: Size of the bubble represents the number of scientists applying.

grantees) and 2186 PoC proposals, using an automated machine learning approach.

Research has indicated that male and senior scientists are more likely to engage in translational research (e.g., Haeussler and Colyvas, 2011; Perkmann et al., 2013). ERC grantees show a similar tendency. Male grantees are more likely to apply for a PoC grant, but there is no difference between male and female peers in the likelihood of winning one. This result can be relevant for policy makers pursuing gender equality in research funding and in light of studies which found a bias against female candidates, such as Schiffbaenker et al. (2022). Our results show that the gender gap does not necessarily emerge in the selection stage, but it can also emerge in the application stage. Thus, policy makers might consider ways to encourage female applicants to equalize the playing field and likely the outcomes.

Early career grantees (i.e., StG) are less likely to apply than senior grantees (AdG), but more likely than mid-career grantees (CoG) to do so. This evidence partly corroborates Munari and Toschi's (2021) claim that the PoC funding scheme may fill a motivational gap of early career scientists, who are often reluctant to embark on research valorization because their priority is to increase their scientific prestige. One potential way to incentivize applications from early career and female candidates is to consider project applications as relevant for tenure and promotion in case of a positive evaluation, regardless of whether the application is successful or not.

Grantees from different countries have similar chances of winning. However, there are major country variations in the grantees' likelihood of applying, resulting in a very different geographical distribution of PoC grants compared to ERC grants. For example, the UK, Germany, France, and Switzerland together won three times more ERC grants between 2007 and 2016 than the Mediterranean countries of Italy, Israel, Greece, and Spain (56 % vs. 17 %). In contrast, they won only one and a half times more PoC grants between 2010 and 2017 (45 % vs. 28 %). This pattern may look surprising, given the strong tradition in applied research and academic entrepreneurship of the first group of countries. However, it is precisely due to this tradition that scientists in such systems often have funding opportunities for translational research from other public and private sources. In contrast, the PoC program meets the strong demand for funding support in the second group of countries, which may lack such opportunities. Indeed, we found that ERC grantees from countries with a higher level of financial support for innovation are less likely to apply for a PoC grant. Policy makers pursuing a more equitable distribution of funds across member countries, might consider cooperating with state-level granting agencies that provide the seed money for larger European applications. Such schemes are already available in countries such as Norway.

ERC grantees from countries with limited levels of collaboration in innovation are much less likely to apply. The PoC supports only the early phases of the valorization process. However, commercialization is a complex process, requiring a receptive environment, with frequent and strong interactions between universities and industry (Gerbin and Drnovsek, 2016). A grantee is unlikely to apply if s/he cannot expect such conditions after the conclusion of the PoC period. We also discovered that ERC grantees from countries with a very high cost of personnel are less likely to apply for an ERC proof-of-concept grant – arguably because in such countries the sum of the grant is not sufficient to fund a doctoral scholarship or to hire a postdoc for a sufficient time.

There were also major variations in the likelihood of applying based on the particular discipline. Those from fields that are more likely to apply for a grant are also more likely to win.

In sum, we argued that studying the effect of a grant on beneficiaries is not sufficient to assess the ability of a funding program to meet its objectives, because of an "iceberg effect." Beneficiaries may represent only a fraction of those in need of or asking for support. The analysis revealed major differences between potential and actual beneficiaries, due to self-selection and selection processes in the application and the evaluation phases, respectively. It reveals how the decision to apply is affected by the interaction between the characteristics of the PoC funding scheme and those of the ERC grantee and his/her context.

When designing the specific characteristics and goals of a research funding program, policy makers should be aware of the indirect impact that their choices have on who applies and, ultimately, on the outcome of the program. Thus, we offer some reflections and suggestions for future research on design choices and tools that can address some possible unintended effects.

First, in the case of the PoC, the amount of the financial contribution of the grant is relatively small and incentivizes the participation of grantees from countries and disciplines with lower research costs and fewer opportunities for alternative funding. On one hand, this is a desirable effect, because it allocates money where it is most needed, to those scientists whose research is more likely to fall into the "valley of death" of innovation. On the other hand, the sum is not sufficiently attractive for some scientists who need larger sums of money. As a result, potentially valuable applications may be lost. Future studies should explore the circumstances in which ERC grantees seek and find support for translational research from other sources, and when they are unsuccessful in doing so. Such studies could help identify methods of dealing with this potential gap. For example, given the differences among European countries and scientific fields in terms of the costs of and need for research personnel and instrumentation, a "one sum fits all" approach may be unsuitable to ensure fair and efficient support for valorization. One solution could be to allow scientists to apply for two or three different amounts of funding – according to their financial needs – and divide the overall budget between the two streams in proportion to the number of applicants.

Second, the support given only for the early phases of commercialization apparently discourages applications from those working in situations where there is little collaboration in innovation. Future research should delve deeper into the factors discouraging translational research in such situations, with the goal of developing complementary tools to avoid losing valuable ideas. For example, PoC grantees from such countries could be given the opportunity to apply for a second grant supporting the middle and later stages of commercialization, and networking tools could be designed to connect them with potential partners and business angels.

The launch of the Horizon Europe framework program in 2021 introduced the European Innovation Council (EIC) as a new major player in supporting breakthrough technologies and game changing innovations. This instrument can potentially fill the gaps identified earlier as it will allow for the award of larger grants at all levels of Technology Readiness. Specifically, one of its funding schemes, the EIC Transition, is restricted to, among others, ERC PoC grantees. Future research on the pool of applicants and grantees of EIC projects would allow for a better understanding of the profiles of those inclined to explore these alternative funding routes.

## Disclaimer

All views expressed in this article are strictly those of the authors and may in no circumstances be regarded as an official position of the European Research Executive Agency, the European Research Council, or the European Commission.

## Declaration of competing interest

None.

## Data availability

The data that has been used is confidential.

## References

Abdoul, H., Perrey, C., Amiel, P., Tubach, F., Gottot, S., Durand-Zaleski, I., Alberti, C., 2012. Peer review of grant applications: criteria used and qualitative study of reviewer practices. PLoS ONE 7 (9), 1–15.

Albarrán, P., Crespo, J.A., Ortuño, I., Ruiz-Castillo, J., 2010. A comparison of the scientific performance of the U.S. and the European Union at the turn of the 21st century. Scientometrics 85, 329–344.

Albers, C.J., 2015. Dutch research funding, gender bias, and Simpson's paradox. Proc. Natl. Acad. Sci. 112 (50), E6828–E6829.

Alon, I., Qi, M., Sadowski, R.J., 2001. Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. J. Retail. Consum. Serv. 8 (3), 147–156.

Alon, I., Boulanger, M., Misati, E., Madanoglu, M., 2015. Are the parents to blame? Predicting franchisee failure. Competitiveness Review 25 (2), 205–217. https://doi.org/10.1108/CR-10-2014-0034.

Auerswald, P.E., Branscomb, L.M., 2003. Valleys of death and Darwinian seas: financing the invention to innovation transition in the United States. J. Technol. Transf. 28 (3), 227–239.

Azoulay, P., Ding, W., Stuart, T., 2007. The determinants of faculty patenting behaviour: demographics or opportunities? J. Econ. Behav. Organ. 63 (4), 599–623.
Blondel, Mathieu, Seki, Kazuhiro, Uehara, Kuniaki, 2013. Block coordinate descent algorithms for large-scale sparse multiclass classification. Mach. Learn. 93 (1), 31–52.
Bol, T., de Vaan, M., van de Rijt, A., 2018. The Matthew effect in science funding. Proc. Natl. Acad. Sci. 115 (19), 4887–4890.
Bornmann, L., 2011. Scientific peer review. Annu. Rev. Inf. Sci. Technol. 45, 199–245.
Boudreau, K., Guinan, E.C., Lakhani, K.R., Riedl, C., 2012. The novelty paradox & bias for normal science: evidence from randomized medical grant proposal evaluations. Harvard Business School Working Paper Series, 13-053.
Boyack, K.W., Smith, C., Klavans, R., 2018. Toward predicting research proposal success. Scientometrics 114, 449–461.
Bradley, S.R., Hayter, C.S., Link, A.N., 2013. Proof of concept centers in the United States: an exploratory look. J. Technol. Transf. 38 (4), 349–381.
Bromham, L., Dinnage, R., Hua, X., 2016. Interdisciplinary research has consistently lower funding success. Nature 534, 684–687.
Bruneel, J., d'Este, P., Salter, A., 2010. Investigating the factors that diminish the barriers to university–industry collaboration. Res. Policy 39 (7), 858–868.
Chang, Chih-Chung, Lin, Chih-Jen, 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 (3), 27.
Clark, B.R., 1998. Creating entrepreneurial universities: organizational pathways of transformation. In: Issues in Higher Education. Elsevier Science Regional Sales, 665 Avenue of the Americas, New York, NY 10010.
Cole, S., Cole, J.R., Simon, G.A., 1981. Chance and consensus in peer review. Science 214 (4523), 881–886.
D'Este, P., Patel, P., 2007. University–industry linkages in the UK: what are the factors underlying the variety of interactions with industry? Res. Policy 36 (9), 1295–1313.
Ding, W.W., Stuart, T.E., 2006. When do scientists become entrepreneurs? The social structural antecedents of commercial activity in the academic life sciences. Am. J. Sociol. 112 (1), 97–144.
Doornenbal, B.M., Spisak, B.R., van der Laken, P.A., 2021. Opening the black box: uncovering the leader trait paradigm through machine learning. Leadersh. Q. 101515.
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'. Res. Policy 35, 1450–1464.
Enger, S.G., 2018. Closed clubs: network centrality and participation in Horizon 2020. Sci. Public Policy 45 (6), 884–896.
Enger, S.G., Castellacci, F., 2016. Who gets Horizon 2020 research grants? Propensity to apply and probability to succeed in a two-step analysis. Scientometrics 109, 1611–1638.
European Commission, 1994. The European report on science and technology indicators 1994. Office for official publications of the European Communities, Luxembourg, ISBN 92-826-9010-5.
European Commission, 1995. Green paper on innovation. COM (95)688 final. Brussels. Accessed at: https://europa.eu/documents/comm/green_papers/pdf/com95_688_en.pdf.
European Commission, 2015. Comparative scientometric assessment of the results of ERC funded projects. In: Bibliometric Assessment Report EC: Directorate-General for Research and Innovation.
European Commission, 2020. European Innovation Scoreboard 2020. Available at: https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_1150.
European Commission, 2021. She Figures 2021. available at:. Publications Office, Luxembourg https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/67d5a207-4da1-11ec-91ac-01aa75ed71a1.
European Council, 2000. Presidency Conclusions. 23-24 March 2000. Accessed at: https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/00100-r1.en0.htm.
Fisher, A., Rudin, C., Dominici, F., 2019. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. J. Mach. Learn. Res. 20 (177), 1–81.
Friedman, Jerome H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Gabrilovich, E., Markovitch, S., 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In: IJcAI, vol. 7, pp. 1606–1611.
García-Quevedo, J., Segarra-Blasco, A., Teruel, M., 2018. Financial constraints and the failure of innovation projects. Technol. Forecast. Soc. Chang. 127, 127–140.
Gerbin, A., Drnovsek, M., 2016. Determinants and public policy implications of academic-industry knowledge transfer in life sciences: a review and a conceptual framework. J. Technol. Transf. 41 (5), 979–1076.
Geuna, A., 1998. Determinants of university participation in EU-funded R&D cooperative projects. Res. Policy 26 (6), 677–687.
Giuri, P., Munari, F., Scandura, A., Toschi, L., 2019. The strategic orientation of universities in knowledge transfer activities. Technol. Forecast. Soc. Chang. 138, 261–278.
Gulbrandsen, M., Smeby, J.C., 2005. Industry funding and university professors'research performance. Res. Policy 34, 932–950.
Haeussler, C., Colyvas, J.A., 2011. Breaking the ivory tower: academic entrepreneurship in the life sciences in UK and Germany. Res. Policy 40 (1), 41–54.
Henriques, L., Schoen, A., Pontikakis, D., 2009. Europe's top research universities in FP6: scope and drivers of participation. European Commission: JRC Technical Notes vol. 53681.
Hoenig, Barbara, 2017. Europe's new scientific elite. Social mechanisms of science in the European research area. Routledge, London & New York.
King, D.A., 2004. The scientific impact of nations. Nature 430, 311–316.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York, NY.
Lam, A., 2011. What motivates academic scientists to engage in research commercialization: 'gold', 'ribbon'or 'puzzle'? Res. Policy 40 (10), 1354–1368.
Larsen, K.R., Becker, D.S., 2021. Automated Machine Learning for Business. Oxford University Press.
Laudel, G., 2006. The art of getting funded: how scientists adapt to their funding conditions. Sci. Public Policy 33 (7), 489–504.
Lee, C.J., Sugimoto, C.R., Zhang, G., Cronin, B., 2013. Bias in peer review. J. Am. Soc. Inf. Sci. Technol. 64 (1), 2–17.
Lepori, B., Veglio, V., Heller-Schuh, B., Scherngell, T., Barber, M., 2015. Participations to European Framework Programs of higher education institutions and their association with organizational characteristics. Scientometrics 105 (3), 2149–2178.
Lockett, A., Wright, M., 2005. Resources, capabilities, risk capital and the creation of university spin-out companies. Res. Policy 34 (7), 1043–1057.
Mantere, S., Ketokivi, M., 2013. Reasoning in organization science. Acad. Manag. Rev. 38 (1), 70–89.
Marsh, H.W., Jayasinghe, U.W., Bond, N.W., 2008. Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. Am. Psychol. 63 (3), 160.
Mayo, N.E., Brophy, J., Goldberg, M.S., Klein, M.B., Miller, S., Platt, R.W., Ritchie, J., 2006. Peering at peer review revealed high degree of chance associated with funding of grant applications. J. Clin. Epidemiol. 59 (8), 842–848.
Meng, Y., 2016. Collaboration patterns and patenting: exploring gender distinctions. Res. Policy 45 (1), 56–67.
Munari, F., Toschi, L., 2021. The impact of public funding on science valorisation: an analysis of the ERC Proof-of-Concept Programme. Res. Policy 50 (6), 104211.
Munari, F., Sobrero, M., Toschi, L., 2018. The university as a venture capitalist? Gap funding instruments for technology transfer. Technol. Forecast. Soc. Chang. 127, 70–84.
Munim, Z.H., Schramm, H.J., 2021. Forecasting container freight rates for major trade routes: a comparison of artificial neural networks and conventional models. Marit. Econ.Logist. 23 (2), 310–327.
Munim, Z.H., Shakil, M.H., Alon, I., 2019. Next-day bitcoin price forecast. J.Risk Financ. Manag. 12 (2), 103.
Murray, F., Graham, L., 2007. Buying science and selling science: gender differences in the market for commercial science. Ind.Corp.Chang. 16 (4), 657–689.
Murray, D.L., Morris, D., Lavoie, C., Leavitt, P.R., MacIsaac, H., Masson, M.E., Villard, M. A., 2016. Bias in research grant evaluation has dire consequences for small universities. PloS One 11 (6), e0155876.
Mutz, R., Bornmann, L., Daniel, H.-D., 2012. Does Gender Matter in Grant Peer Review? An Empirical Investigation Using the Example of the Austrian Science Fund. Z. Psychol. 220 (2), 121–129.
Neufeld, J., Huber, N., Wegner, A., 2013. Peer review-based selection decisions in individual research funding, applicants'publication strategies and performance: the case of the ERC Starting Grants. Res.Eval. 22 (4), 237–247.
O'Gorman, C., Byrne, O., Pandya, D., 2008. How scientists commercialise new knowledge via entrepreneurship. J. Technol. Transf. 33 (1), 24–43.
Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A., Sobrero, M., 2013. Academic engagement and commercialisation: a review of the literature on university–industry relations. Res. Policy 42 (2), 423–442.
Piro, F.N., Børing, P., Scordato, L., Aksnes, D.W., 2020. University characteristics and probabilities for funding of proposals in the European Framework Programs. Sci. Public Policy 47 (4), 581–593.
Rasmussen, E., 2008. Government instruments to support the commercialisation of university research: lessons from Canada. Technovation 28, 506–517.
Rasmussen, E., Sørheim, R., 2012. How governments seek to bridge the financing gap for university spin-offs: proof-of-concept, pre-seed, and seed funding. Technol.Anal. Strateg.Manag. 24 (7), 663–678.
Research Council of Norway, 2021. Science & Technology Indicators for Norway 2021. The Research Council of Norway, Oslo.
Rodríguez-Navarro, A., Narin, F., 2018. European paradox or delusion—are European science and economy outdated? Sci. Public Policy 45 (1), 14–23.
Sandström, U., Hällsten, M., 2008. Persistent nepotism in peer-review. Scientometrics 74, 175–189.
Schiffbaenker, H., van den Besselaar, P., Holzinger, F., Mom, C., Vinkenburg, C., 2022. Gender bias in peer review panels:–"The elephant in the room" 1. available at:. In: Inequalities and the Paradigm of Excellence in Academia. Routledge, pp. 109–128.
Seeber, M., 2013. Efficacy and limitations of research steering in different disciplines. Stud. High. Educ. 38 (1), 20–38.
Seeber, M., Vlegels, J., Reimink, E., Marušić, A., Pina, D.G., 2021. Does reviewing experience reduce disagreement in proposals evaluation? Insights from Marie Skłodowska-Curie and COST Actions. Res. Eval. 30 (3), 349–360.
Seeber, M., Vlegels, J., Cattaneo, M., 2022. Conditions that do or do not disadvantage interdisciplinary research proposals in project evaluation. J. Assoc. Inf. Sci. Technol. 73 (8), 2330–1635.
Sellenthin, M.O., 2009. Technology transfer offices and university patenting in Sweden and Germany. J.Technl.Transf. 34, 603–620.
Stephan, P., Gurmu, S., Sumell, A., Black, G., 2007. Who's patenting in the university? Evidence from the suvery of doctorate recipients. Econ. Innov. New Technol. 16 (2), 71–99.
Tamblyn, R., Girard, N., Qian, C.J., Hanley, J., 2018. Assessment of potential bias in research grant peer review in Canada. CMAJ 190, E489–E499.
Tartari, V., Salter, A., 2015. The engagement gap: exploring gender differences in university-industry collaboration activities. Res. Policy 44 (6), 1176–1191.

Thune, T., Reymert, I., Gulbrandsen, M., Aamodt, P.O., 2016. Universities and external engagement activities: particular profiles for particular universities? Sci. Public Policy 43 (6), 774–7786.
Truong, A., Walters, A., Goodsitt, J., Bruss, C.B., Farivar, R., 2019. Towards automated machine learning: evaluation and comparison of AutoML approaches and tools. In: IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1471–1479. https://doi.org/10.1109/ICTAI.2019.00209.
Van den Besselaar, P., Leydesdorff, L., 2009. Past performance, peer review and project selection: a case study in the social and behavioural sciences. Res.Eval. 18 (4), 273–288.
Van den Besselaar, P., Sandström, U., Schiffbaenker, H., 2018. Studying grant decision-making: a linguistic analysis of review reports. Scientometrics 117, 313–329.
Vilkkumaa, E., Salo, A., Liesiö, J., Siddiqui, A., 2015. Fostering breakthrough technologies—how do optimal funding decisions depend on evaluation accuracy? Technol. Forecast. Soc. Chang. 96, 173–190.
Volker, B., Steenbeek, W., 2015. No evidence that gender contributes to personal research funding success in The Netherlands: a reaction to van der Lee and Ellemers. Proc. Natl. Acad. Sci. 112 (51), E7036–E7037.
Wanzenböck, I., Lata, R., Ince, D., 2020. Proposal success in Horizon 2020: a study of the influence of consortium characteristics. Quant.Sci.Stud. 1 (3), 1136–1158.

**Marco Seeber**, Ph.D., is Associate Professor at the Department of Political Science and Management of the University of Agder (Norway). He holds a Ph.D. degree in organizational studies, and previously worked at CNR (Italy), University of Lugano (Switzerland) and Ghent University (Belgium). His works examines the governance of Higher Education systems and universities, research policy and research evaluation. His works appeared, among others, on Research Policy, Higher Education, Studies in Higher Education, Research Evaluation, JASIST, Public Management Review. He is the Co-Editor in Chief of the European Journal of Higher Education.

**Ilan Alon** (Ph.D., Kent State University, USA) is Professor of Business and Economics in Ariel University (Israel) and Agder University (Norway). A prolific author, he is also the editor-in-chief of the International Journal of Emerging Markets and European Journal of International Management. Previously, Alon has held research positions at Harvard University (USA), Georgetown University (USA), Copenhagen Business School (Denmark), and University of International Business and Economics (China). Alon has worked with government bodies (e.g., USAID), non-profit organizations (e.g., Word Trade Organization), multinational companies (e.g., illy and Disney) and international associations (Chinese Globalization Association) on various projects, ranging from capacity development at the macro level to international business development at the firm or project level.

**David G. Pina**, Ph.D. is research program officer at the European Research Executive Agency. He is a biochemist by training. After graduating in 1999 from the University of Coimbra, Portugal, he studied the structural stability of proteins and completed his PhD in 2004 at the University of Salamanca, Spain. He worked as a postdoctoral fellow at the Curie Institute in Paris and the University of Cambridge, UK, before joining the European Commission in 2007 as a project officer. He has since then been involved in the management of several EU's research programmes, such as the Marie Curie Actions or the "Widening Participation". His works on grant evaluation appeared in Elife, Research Evaluation and PlosOne.

**Fredrik Niclas Piro**, Ph.D., is Research Professor at the Nordic Institute for studies in Innovation, Education and Research (NIFU), Norway. He holds a Ph.D. degree in public health medicine. His main fields of interests are bibliometrics, research funding and collaboration/competition in the European Framework programs. His scientific papers have been published in e.g., Research Evaluation, Science and Public Policy, Scientometrics and Journal of the American Society for Information Science and Technology (JASIST).

**Michele Seeber**, Ph.D., is a research technician at the Life Science Department of the University of Modena and Reggio Emila (UniMoRe, IT). He holds a PhD in Chemistry, and he previously worked at the Physics, Informatics and Mathematics Department of UniMoRe, and at the Biochemistry Department of the University of Zurich (CH). His main research interest is the development of methods, software and procedures for analysis of molecular dynamics simulations, docking and screening experiments, and bibliometric studies.