An empirical assessment of the impact of public research contracts on scientific productivity*

César Alonso-Borrego† — Universidad Carlos III

Antonio Romero-Medina — Universidad Carlos III

Rocío Sánchez-Mangas — Universidad Autónoma de Madrid

Abstract

We study the effect of a Program, aimed at financing high quality researchers to integrate them into the Spanish scientific system, on the relative ex post performance of the researchers awarded. We assess the effect of the contract status on the scientific productivity of applicants, in the four-year period after application. Both the conditional regression and the matching results show that the contract status has no effect on the average number of published contributions, but it exhibits, for several areas, a positive effect on the scientific performance (as measured by the impact of contributions). This result points out the success of the Program in increasing the scientific impact of the Spanish system.

Keywords: Brain Gain, Research Productivity, Government Research Programs, Human Capital, Policy Evaluation.

JEL Classification Numbers: O38, D78, C21, I23, O31.

*We thank the Subdirección General de Formación y Movilidad del Personal Investigador (Spanish Office of Training and Mobility of Researchers) for providing the raw data and Miguel Benavente for help in data handling. Research funding from Ministry of Education, Grant Nos. ECO2009-11165 and ECO2012-31358 (Alonso-Borrego), ECO2011-25330 (Romero-Medina) and ECO2009-10287 (Sánchez-Mangas) is acknowledged. We thank comments from audience in XXXVII SAEe, Vigo, December 2012.

†Corresponding Author. *Department of Economics, Universidad Carlos III de Madrid*, 28903 Getafe, Spain. e-mail: cesar.alonso@uc3m.es.

1 Introduction

This paper studies if the Ramón y Cajal (R&C) contract has a differential impact on the research output of the researchers awarded with Ramón y Cajal contracts years later. For such purpose, we exploit data on applications in several calls of the Program, as well as individual and curricular information of the applicants.

The Ramón y Cajal Program was introduced in 2001 by the Spanish Government. It is an ambitious publicly funded program aimed at providing career paths to high-quality researchers and to integrate them within the national scientific system. It was created in the general context of a lack of R&D personnel in Spain and with Spanish Universities hiring policies being called into question.

The researchers hired under the program benefitted from a well-defined career path, with a 5-year contract in a Spanish research centers, and the possibility to join permanent research positions at the end of the contract (see Sanz Menéndez et al, 2002; Sanz Menéndez, 2003).

The selection procedure was centralized in an evaluation agency, Agencia Nacional de Evaluación y Prospectiva (ANEP), which appraised all eligible applicants based on predetermined rigorous and objective evaluation criteria based mainly on the candidates' scientific record. For this purpose, 24 evaluation committees of national and international experts, one for each research field, were constituted by the evaluation agency.[1] The selection, centralized and external to the research centers, was a novel feature in the public policy design aimed at hiring of new researchers. The Government was aware of the pervasive tradition of inbreeding in the Spanish scientific system, and its negative consequences on scientific performance (Eisenberg and Wells, 2000; Soler, 2001). Four years after each call, the performance of each granted applicant during the contractual period was evaluated. A positive evaluation implies the possibility to receive a new contract that facilitates her access to a tenured contract in the research center.

Our findings confirm that the assessment process was based on the applicant's research

[1] A list of the 24 research areas is shown in the Appendix, Table A1.

C.V. and that the available curricular information mattered for the grading that the assessment committees gave to each applicant. In order to assess the effectiveness of the Program, we analyze the impact of the contract status on researchers' scientific productivity four years after the call. To circumvent potential selection biases due to differences between successful and unsuccessful applicants, we control for observed curricular characteristics that yield a similar probability of contract at the time of the call. Our main results show that the contract had no impact on the number of published contributions, but had a positive effect on the quality of contributions.

The rest of the paper is organized as follows. In section 2, we introduce the main data set of applications, and the complementary data set on the applicants' curricular information and preliminary results. In section 3, we evaluate the effectiveness of the Program in the scientific productivity of successful applicants. In section 4, we summarize the major results and discuss their policy implications, and conclude.

2 Data and preliminary evidence

The main data set, provided by the Dirección General de Investigación of the Spanish Ministry of Education, records all applications in the first seven calls of the program, from 2001 to 2007. We excluded observations with missing values for individual characteristics, which represent less than one percent of all observations. Each applicant information includes her research area, the institution and the year when she earned her PhD, her country of residence and nationality, as well as the score received in the assessment process and whether she was granted a contract.

In Table 1, we provide the distribution of all the applicants and the distribution of successful applicants. The distribution is not uniform in time, as it is not the number of contracts offered, which decreased substantially since 2003. In addition, since 2004, eligibility was restricted to earning the PhD in the last 10 years, and a minimum 2-years postdoctoral stay in a center different than that in which the Ph.D. was obtained.

We observe that applications are dominated by men. The research areas of Biology, Chemistry and Medicine cope about 60% of the applicants. It must be noted that the gender distribution is strongly unequal across research areas. Physics and Engineering are strongly dominated by men, amount 80% of applicants. In Chemistry and Business, men represent about 60%. However, in Social Sciences and in Biology, the proportion of men is around 52%. Medicine is dominated by women, with 54% of applicants. With respect to time elapsed since the PhD, the majority of applicants earned their Ph.D. within 3 to 6 years before the call.

The curricular information has been collected from a complementary data source, the free net resource Publish or Perish (Harzing, 2007). Publish or Perish retrieves academic contributions by author using the Google Scholar database, which provides the title, the source, the year and the authors of the contribution. Google Scholar is generally praised for its speed (Bosman et al., 2006) and its high correlation with alternative bibliometric sources (See Harzing, 2012, and Harzing and van der Wal, 2011, for a comparison of citation analysis using different data sources). Whenever the contribution was published in a scientific journal, the journal information is also reported. For each applicant, we measure her number of distinct contributions and, among these, the number of published papers. In order to weight the quality of each contribution, we use the Journal of Citation Reports (JCR), which provides the impact factors of the international journals listed in its database. The impact factor of a journal is calculated on the basis of the average number of citations attained by the contributions published in that journal. We use the JCR impact factors to measure both the quality of each candidate, as well as the quality of the center where each candidate earned her PhD, defined as the average number of citations to all the works published in JCR journals by all the researchers affiliated to the center. The curricular information is updated up to 2007.

We concentrate on the first calls, until 2003, as we have to exploit curricular information several years after the call. Data on 2001 is also disregarded, given the special characteristics of the first call, which might harden the comparability of applications with subsequent calls (Alonso-Borrego et al, 2013). Our final sample, therefore, contains 4,967 applicants between 2002 and 2003.

We use three measures of scientific quality of each applicant: her number of contributions listed in the JCR database, the average impact factor of her JCR publications, and the maximum impact factor among the JCR journals in which she has published. The two impact factor measures are based on the corresponding impact of the journal in which she has published each contribution.

In Table 2 we summarize the curricular information of applicants by contract status. Besides, we break down the sample by applicants' characteristics: gender, research area, and time elapsed since the PhD. For all categories considered, we observe that, at the time of the call, granted researchers have, on average, more published contributions and a higher scientific impact (either average or maximum impact) than non-granted researchers. Nevertheless, given the high standard deviations, most differences are not significant. We also find that the three measures of scientific quality differ substantially by area, reflecting differences in the usual number of papers and citations among areas.

To ascertain the factors that are relevant for the committees' assessment of applicants, we consider a conditional analysis of the applicants' scores on their individual and curricular characteristics. The OLS estimation results are shown in Table 3. The first column reports the full sample regression of score on applicant's characteristics, using qualitative variables to allow for differences among areas. We also report separate estimations for each of the Publish or Perish areas.[2] In all the estimates, both for the full sample and by areas, a high proportion of the variance of score is explained. In most areas, scientific quality of applicants prove to be determinant in the committees' assessments, and in general the quality of contributions matters more than its quantity. Furthermore, the quality of the center in which the Ph.D. was obtained is also a relevant factor in Physics, Economics and Humanities. In the case of Economics and Humanities, the curricular variables are non significant. The small sample size in Economics lead to very imprecise estimated coefficients. This is not the case, though, of Humanities, what suggests that the committee's assessment relies on different criteria. In Physics, the number of JCR papers appears more relevant than the quality of contributions. It is also interesting that PhD tenure has an inverted-U

[2] A list of the 7 areas reported in Publish or Perish is shown in the Appendix, Table A2.

effect on applicants' score in most areas. Given that we are also controlling for scientific quality, this variable, together with the quality of the PhD center, might be capturing other unobserved quality features. For instance: papers on under revision; forthcoming papers (but not published at the year of the call); the quality of the research agenda of the candidate, etc.

Among the individual characteristics, we find a positive and significant gender effect in favor of men, which ranges between 3 and 5 percentage points. This result suggests that men are slightly better graded than women with similar scientific quality, pointing out a certain degree of gender discrimination. Also, we have included the quality of the PhD institution, measured by the cumulated impact of the contributions listed in the JCR of its faculty members, which has a positive effect, and it is significant in several areas.

3 Empirical approach

We are mostly intrigued about the impact of the contract status on the ex-post performance of researchers. For that purpose, we consider the scientific outputs of applicants in the four years after the call. Given the data constrains, we consider the time horizon chosen to be sufficient to test the potential influence of the contract. It is, though, consistent with the usual time span for the tenure decision taken by the research centers. Moreover, such time span seems to be in coherence with the maximum time length needed to undertake a peer-reviewed publication process of scientific contributions.

Our relevant policy variable is a binary variable indicating whether the individual was granted a Ramón y Cajal contract, that we denote as D_i, which takes on value 1 if the researcher i has been granted a contract and zero otherwise. Our concern is whether the contract status affects the reseacher's productivity outcome Y_i in the four-year period after the call. We undertake the analysis using three alternative outcome variables, measuring scientific performance of researchers. These variables are the number of contributions published in journals listed in the JCR, the average impact of such contributions, and the maximum impact factor among the JCR journals in which she has published.

As it is well known, the ideal evaluation problem, for a given researcher, consists on comparing her two potential outcomes depending on whether she had and she had not a contract, denoted as Y_{0i} and Y_{1i}, respectively. If both counterfactual outcomes were observed for researcher i, the impact of the contract for such researcher would simply be $(Y_{1i} - Y_{0i})$, and we then could calculate the average treatment effect, i.e., the average impact of the contract computing the sample counterpart of $E\left(Y_{1i} - Y_{0i}\right)$, where $E\left(.\right)$ denotes the mean operator (see Rosenbaum and Rubin, 1983).

Since having or not having a contract are mutually exclusive, for each researcher we just observe either $D_i = 1$ or $D_i = 0$, and therefore we just observe her outcome under one of the two situations, i.e.,

$$Y_i = Y_{0i} + \left(Y_{1i} - Y_{0i}\right) D_i. \tag{1}$$

Assuming that the effect of the contract is homogeneous, so that $Y_{1i} - Y_{0i} = \rho$, and denoting $\alpha = E\left(Y_{0i}\right)$, we can write the expression above as,

$$Y_i = \alpha + \rho D_i + u_i, \tag{2}$$

where u_i is an error term capturing unobserved individual differences in scientific productivity. Given that for each researcher we just observe one potential outcome, the regression based on observed outcomes would provide the mean differences in outcomes between researchers granted with a contract and researchers without a contract,

$$E\left(Y_i \middle| D_i = 1\right) - E\left(Y_i \middle| D_i = 0\right). \tag{3}$$

which, unless contract status were purely random, will differ from ρ. We know, indeed, that contract status depends on researchers' characteristics, so that researchers' potential outcomes Y_{1i}, Y_{0i} are not independent of the contract status D_i. Essentially, those researchers with a contract are likely to be more productive than researchers without a contract anyway. As a consequence, if we consider the naive regression (2) to estimate ρ as the mean difference in productivity between successful and non-succesful applicants, such estimate will be contaminated by selection bias, since $E\left(u_i \middle| D_i\right) \neq 0$. Presumably, the naive mean-difference estimator would exacerbate the positive impact of the contract.

We observe, though, additional individual information $\mathbf{X}_i$, which corresponds to individual curricular information at the time of application. If we assume that, after conditioning on the exogenous covariates in $\mathbf{X}_i$, the potential outcomes are mean-independent of D_i, so that $\mathbf{X}_i$ determines contract status, then

$$E\left(Y_{ji} \mid D_i, \mathbf{X}_i\right) = E\left(Y_{ji} \mid \mathbf{X}_i\right) \qquad j = 0, 1.$$

This conditional mean-independence assumption (also called "selection on observables") implies that we can the causal effect of the contract can be estimated from the augmented specification

$$Y_i = \alpha + \rho D_i + \boldsymbol{\beta}' \mathbf{X}_i + u_i, \tag{4}$$

where now $E\left(u_i \mid D_i, \mathbf{X}_i\right) = 0$.

Consequently, OLS estimation of (4) will yield a consistent estimate of the impact of the contract. This specification establishes that, conditional on $\mathbf{X}_i$, the causal effect of the contract is the same for any applicant and equal to ρ. In particular, the causal effect of the contract for the whole population of applicants (that is usually called the ATE, average treatment effect) coincides with the causal effect of the contract for those who actually earned a contract (usually called the ATT, average treatment effect on the treated). We can allow for differences in the causal effect across different groups in the population, particularly, that ATE and ATT be different, assuming a weaker version of the mean-independence assumption such that

$$E\left(Y_{0i} \mid D_i, \mathbf{X}_i\right) = E\left(Y_{0i} \mid \mathbf{X}_i\right),$$

i.e., conditional on the observed covariates, the outcome under no contract is mean-independent of the contract status. Under such assumption, we must consider a the extended model (see Wooldridge, 2002)

$$Y_i = \alpha + \rho D_i + \boldsymbol{\beta}' \mathbf{X}_i + \boldsymbol{\gamma}' D_i \mathbf{X}_i + u_i, \tag{5}$$

where now the causal effect is equal to $\rho D_i + \boldsymbol{\gamma}' \mathbf{X}_i$, so it varies with the covariates. To calculate the ATE, we must evaluate such expression at $E\left(\mathbf{X}_i\right)$, while to calculate the ATT, it must be evaluated at $E\left(\mathbf{X}_i \mid D_i = 1\right)$.

Alternatively, we can consider matching estimators of the impact of the contract on those researchers who actually obtained a contract, i.e., the average treatment effect on the treated. The idea is to compare the outcomes of granted researchers with the outcomes of selected non-granted researchers that are similar to the first ones except for their contract status. The criterion to match each granted researchers with a non-granted researcher will be the propensity score, i.e., the probability of being granted a contract, conditional on the applicant's characteristics at the time of the call. The validity of our propensity score matching relies on the validity of the Conditional Independence assumption (CIA), i.e., the selection of researchers in the treatment group (granted researchers) or in the control group (non granted researchers) is only based on observables (Rosenbaum and Rubin, 1983). In other words, once we control for these observables (CV quality and other individual characteristics at the time of the call), being granted or not cannot depend on output. Taking into account that our performance measure consists of the CV quality four years after the call, we are quite confident that the CIA holds.

4 The performance of Ramón y Cajal researchers

In this section we report the alternative estimates of the causal effect of the contract based on alternative assumptions. The covariates that we consider in the empirical analysis contain curricular information at the time of the call: the researcher curricular information, the time elapsed since the researcher earner her PhD, the quality of her PhD institution, and the researcher's gender. The validity of the regression estimate of the causal effect in (4) relies on absence of unobserved differences between granted and non-granted researchers that affect their potential outcomes.

In Table 4, we report OLS linear regression estimates for the full sample of applicants for the three alternative outcomes. For the sake of camparison, we report the naive unconditional estimates of the impact of the contract corresponding to eq. (1), and the conditional estimations of the impact of the contract based on (4). In all estimations we control for differences across scientific areas using a set of area binary dummies. For any of the outcomes, the naive estimates yield a highly positive and significant effect of the contract. It must be noted, though, that there are substantial differences between researchers with and without contract that, on average, make granted researchers more productive than non-granted researchers, so we expect unconditional estimates to be contaminated by a strongly positive selection bias. This is confirmed by the conditional estimates of the impact of the contract, which are much smaller in magnitude. In fact, we find that, when we control for researcher' characteristics at the time of the application, the contract has no significant effect on the number of published contributions in the four-year period after the call, so there are not differences in quantity by contract status. If we regard the influence of the scientific contributions, measured by either the average impact or the maximum inpact, we find that contract status entails significantly positive differences in quality. This results suggest that researchers keep producing scientific output at a similar pace irrespective of their contract status, but researchers with a contract achieve a higher scientific influence.

When we disaggregate by areas, as shown in Table 5, we get similar results than with the full sample. Again, the naive estimates of the causal effects, reported in Table 5, are positive and significant, both for quantity and quality of scientific contributions, in most of the areas. The mos tinteresting results correspond with the conditional estimates, using pre-contract researchers' characteristics as covariates. We do not found differences in the quantity of scientific contributions between granted and non granted applicants. But some differences across areas arise in the scientific quality. In particular, in the case of Biology, Chemistry and Physics, we find a positive effect of the contract on the average impact and the maximum impact. Most the effects are not significant for Economics, Engineering, Medicine and Humanities. In the case of Economics, the sample size is very small, and the causal effect is only significant for the shift in the impact distribution in accordance with the best researcher's paper.

In the case of Engineering, we find positive but small effects in the average and the maximum impact. However, there can exist differences in the scientific standards that rule this discipline: the development of patents is particularly relevant in many Engineering fields, as much as the published contributions in scientific journals.

We believe that the absence of a positive causal effect in Medicine is an intriguing result, which deserves further investigation. It suggests that non-granted researchers keep their research career in an environment that favors their scientific impact. Issues like the availability and quality of laboratories and other available resources can be behind this.

The estimation results for the propensity score matching estimates are shown in Tables 6 and 7. Granted applicants are then matched with non-granted applicants who are similar in their propensity score. To check whether results might be influenced by the 'similarity' criterion, we consider three alternative criteria: kernel, nearest neighbor and stratification. We have used the Stata procedure written by Becker and Ichino (2002). The variables used to estimate the probability of being granted a contract are the same that we have used as covariates in the conditional regression estimates.[3] Essentially, alternative matching criteria do not differ except for the significance of the estimated effects: kernel and stratification methods yield very similar results, and nearest neighbor method is typically less precise. Qualitatively, the results resemble the obtained for the conditional regression estimates. Again, the contract status does not affect the quantity of publications, and the contract has a positive impact in our two measures of scientific quality for Biology, Chemistry and Physics. Interestingly, contract status for Medicine does not yield significant differences in scientific quality.

In general, the results suggest that, for researchers who are comparable in their ex ante characteristics, the contract status does not yield differences in the number of scientific contributions four years after the call. In addition, researchers with a Ramón y Cajal contract show, on average, a scientific impact four years after the call higher than that for comparable researchers without contract. This difference is, indeed, significant for several areas. We find, though, an exception in the case of Medicine.

[3] The estimated propensity score yields predicted probabilities of being granted a contract that hold the balancing property. Such property establishes that, conditioning on the propensity score, the distribution of the explanatory variables is not different for granted and not granted researchers.

5 Conclusions

The Ramón y Cajal Program was created to ameliorate the shortage of funds for research personnel and to improve the quality of the Spanish R&D system. For that purpose, the program provided funding to recruit quality researchers and to provide them an entry point into the R&D system. The Program was successful in selecting high quality researchers, the selection being based on curricular merits.

We have analyzed the effect of the program on the productivity of the selected researchers and compare them with scholars with similar curricular characteristics that were not awarded with a Ramón y Cajal contract. We have undertaken two alternative approaches to estimate the causal effect of the contract: conditional regression and propensity score matching procedures. Overall, the results provided by both methods are alike.

We find that the selection process was based on the applicant's research curriculum and that the researchers maintain, once in the Program, a quantitative level of scientific production comparable with similar researchers that were marginally rejected in the selection process. When we consider the scientific impact of the researchers, we find it is at least as high as that for non-granted researchers, and it is significantly higher in several areas.

In some areas for which we do not find a significant effect of the contract, particularly Engineering and Humanities, there can exist different scientific practices and standards. In such cases, our curricular measures can render insufficient to characterize the research merits of the candidates. Also, we find a differential result for Medicine. According to our conditional regression estimates, the contract has a negative causal effect on scientific impact; the matching estimator is also negative but significant. This is an intriguing result; further investigation would require additional data to ascertain where those non-granted researchers that ex ante are comparable with granted researchers have developed their research career in the years following the program call.

Our results point out the success of the Program in increasing the scientific impact of the Spanish system in several research areas. Interestingly, the program does not appear to have an effect on the quantity of scientific papers produced, but it has on the impact of the scientific contributions. This is an important result, that supports that policies aimed at increasing the stock of human resources in scientific research help to rise the international impact of the Spanish R&D system.

The Program is a very interesting example of oriented public funding to increase research personnel in Spanish research centers. We highlight, among its major features, that candidates' assessment is undertaken by a governmental agency on the basis of their objective scientific merits, without the participation of the centers. This centralized selection ensures that the standards under which granted researchers are chosen are similar among candidates within the same research area. To benefit from the Program, research center must hire researchers only among the pool of granted candidates. Our results confirm the relevance of scientific quality in candidates' assessment. Also, we have found that the scientific quality of the candidates ex ante (at the time of the call) is very relevant to explain their ex post performance. The thorough selection of candidates is behind the successful performance of the Program. We believe that the centralized selection procedure plays a determinant role in the outcome on the Program, given the pervasive tradition of inbreeding in the Spanish scientific system and its negative consequences on scientific performance.

References

[1] Alonso-Borrego, C., Romero-Medina, A. Sánchez-Mangas, and Triossi, M. (2013), "Boosting Scientific Research: Evidence from a Public Program". Revista de Economía Aplicada (forthcoming).

[2] Becker, S., Ichino, A. (2002), "Estimation of average treatment effects based on propensity scores". The Stata Journal, 2(4), 358-377.

[3] Bosman, J, Mourik, I. van, Rasch, M., Sieverts, E. and Verhoeff, H. (2006), "Scopus reviewed and compared. The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar". Utrecht: Utrecht University Library, http://igitur-archive.library.uu.nl/DARLIN/2006-1220-200432/Scopus doorgelicht & vergeleken - translated.pdf.

[4] Eisenberg, T. and M. T. Wells (2000), "Inbreeding in law school hiring: Assessing the performance of faculty hired from within". Journal of Legal Studies, XXIX, 369–388.

[5] Harzing, A.W. (2013), "A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel Prize winners". Scientometrics (fortcoming).

[6] Harzing, A.W. and van der Wal, R. (2011), "Google Scholar as a new source of citation analysis?". Ethics in Science and Environmental Politics 8, 61-73.

[7] Rosenbaum, P. R. and D. B. Rubin (1983), "The central role of the propensity score in observational studies for causal effects". Biometrika 70(1), 41–55.

[8] Sanz Menéndez, L. (2003), "Coping with researchers' labour market problems through public policy: The Spanish Ramón y Cajal Program", in: Avveduto, S. (Ed.), *Fostering the Development of Human Resources for Science and Technology*, Biblink Editori, 241-274.

[9] Sanz Menéndez, L., M. J. Jerez, A. Romero-Medina, I. Marqués and A. Martínez (2002), “Una nueva política de recursos humanos en I+D: el Programa Ramón y Cajal”. *Economía Industrial* 343, 149–160.

[10] Soler, M. (2001), "How inbreeding affects productivity in Europe". Nature 411, 132.

Table 1. Distribution of applicants and contracts
Percentages in each category

ALL							
Applicants[a]	6842						
% Contracts[b]	2224						
By Year (%)	2002	2003	2004	2005	2006	2007	
Applicants[a]	36.6	37.3	19.6	18.7	19.5	21.5	
% Contracts[b]	19.8	27.5	22.1	18.9	18.5	16.4	
By Gender (%)	Fem.	Male					
Applicants[a]	41.5	58.5					
% Contracts[b]	28.1	36.2					
By Area (%)	Biol.	Econ.	Chem.	Eng.	Medic.	Phys.	Hum.
Applicants[a]	35.6	2.4	15.8	11.5	12.2	9.0	13.5
% Contracts[b]	39.4	31.3	44.8	44.5	33.9	42.2	24.3
By Ph.D tenure (%)	Up to 2 yr.	3-6 years	> 6 years				
Applicants[a]	15.1	55.6	29.2				
% Contracts[b]	28.2	33.4	32.1				

[a]The percentages of applicants by category add over 100%, since a fraction of them apply in several years and/or in several areas.
[b]Share of granted researchers in the corresponding category.

Table 2. Curricular information of applicants by contract status

	Average No. of papers		Average Impact factor		Maximum Impact factor	
CONTRACT:	Yes	no	Yes	no	Yes	no
All	2.5	1.6	1.8	1.0	2.4	1.2
	(6.4)	(5.9)	(3.4)	(2.0)	(5.2)	(3.2)
By Gender						
Female	2.6	1.8	2.1	1.1	2.7	1.4
	(6.6)	(6.0)	(3.1)	(2.0)	(5.1)	(3.6)
Male	2.4	1.5	1.7	0.9	2.3	1.1
	(6.3)	(6.1)	(3.7)	(2.0)	(5.4)	(2.9)
By Research area						
Biology	2.7	2.5	2.4	1.4	3.1	1.8
	(6.3)	(8.5)	(4.2)	(2.2)	(5.7)	(4.1)
Economics	1.9	0.9	0.8	0.5	0.9	0.5
	(1.9)	(3.6)	(0.9)	(1.4)	(1.0)	(1.5)
Chemistry	3.0	1.7	1.7	1.0	2.4	1.2
	(5.2)	(3.9)	(2.5)	(1.5)	(4.1)	(2.3)
Engineering	0.9	0.5	0.5	0.2	0.4	0.2
	(2.6)	(1.5)	(1.5)	(0.4)	(1.0)	(0.4)
Medicine	4.7	2.6	2.7	1.8	4.2	2.1
	(9.7)	(6.4)	(4.0)	(2.9)	(8.0)	(4.3)
Physics	2.3	0.8	1.9	0.9	2.1	1.0
	(9.6)	(2.1)	(3.0)	(2.3)	(3.7)	(2.5)
Humanities	0.5	0.2	0.8	0.2	1.2	0.2
	(1.4)	(0.9)	(2.9)	(0.6)	(6.0)	(0.7)
By Ph.D. tenure						
Up to 2 years	0.9	0.6	1.2	0.6	1.3	0.6
	(2.2)	(2.3)	(3.2)	(1.7)	(3.5)	(2.2)
3-6 years	2.1	1.4	2.0	1.0	2.7	1.3
	(4.1)	(5.0)	(3.4)	(2.1)	(5.5)	(3.2)
More than 6 years	3.6	2.7	1.9	1.3	2.7	1.6
	(9.6)	(8.7)	(3.6)	(2.2)	(5.7)	(4.0)

Standard deviation in parentheses below the sample mean of each category.

Table 3. Assessment of candidates

Dependent variable: Score

	ALL	By Area: Biol.	Econ.	Chem.	Eng.	Medic.	Phys.	Hum.
Constant	$58.41^{\S}$	$49.74^{\S}$	$56.66^{\S}$	$46.12^{\S}$	$54.14^{\S}$	$55.31^{\S}$	$68.39^{\S}$	$53.81^{\S}$
	(1.56)	(2.28)	(7.47)	(3.64)	(3.88)	(3.88)	(4.34)	(3.36)
# Papers	0.03	-0.02	1.18	0.11	$1.52^{\dagger}$	0.002	0.13^{*}	1.19
	(0.05)	(0.05)	(1.32)	(0.16)	(0.57)	(0.12)	(0.06)	(1.33)
Avg. IF	$0.91^{\S}$	$0.74^{\S}$	-0.34	$1.96^{\S}$	$1.21^{\dagger}$	$0.62^{\S}$	0.80	$1.04^{\dagger}$
	(0.15)	(0.18)	(3.57)	(0.43)	(0.53)	(0.20)	(0.64)	(0.44)
Gender	$3.54^{\S}$	$4.20^{\S}$	-1.22	$5.05^{\S}$	$5.40^{\dagger}$	$2.73^{\dagger}$	0.02	$4.33^{\dagger}$
	(0.60)	(0.89)	(5.00)	(1.85)	(2.53)	(1.29)	(2.12)	(1.89)
PhD tenure	$1.45^{\S}$	$1.56^{\S}$	-1.66	$2.37^{\S}$	0.95^{*}	$1.61^{\S}$	$1.36^{\dagger}$	$2.33^{\S}$
	(0.18)	(0.34)	(1.37)	(0.44)	(0.50)	(0.60)	(0.54)	(0.40)
PhD tenure^2	$-0.05^{\S}$	$-0.05^{\dagger}$	0.10^{*}	$-0.08^{\S}$	-0.03	$-0.10^{\S}$	$-0.06^{\dagger}$	$-0.07^{\S}$
	(0.01)	(0.02)	(0.05)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)
PhD center	$0.55^{\S}$	$0.61^{\S}$	0.64	$0.55^{\dagger}$	$0.63^{\S}$	0.49^{*}	0.11	0.10
quality	(0.09)	(0.17)	(0.41)	(0.26)	(0.23)	(0.29)	(0.29)	(0.24)
# obs.	4,967	1,995	104	694	394	730	434	616
R^2	0.11	0.08	0.08	0.13	0.13	0.07	0.05	0.08

Binary dummies for research areas (ANEP) included in the full sample regression. Gender takes on value 1 if male and 0 otherwise. Robust standard error in parentheses.
*,† ,§Significant at 10%, 5% and 1% levels, respectively.

Table 4. Regression estimates of causal effect of the contract
Full sample.

Outcome variable	Unconditional ATE		Conditional: Restricted ATE		Unrestricted ATE		ATT	
# papers	$0.79^{\S}$	(0.23)	0.19	(0.26)	0.23	(0.24)	0.01	(0.33)
Avg. impact	$0.64^{\S}$	(0.10)	$0.21^{\dagger}$	(0.08)	$0.19^{\dagger}$	(0.08)	$0.24^{\dagger}$	(0.10)
Max. impact	$0.85^{\S}$	(0.13)	$0.29^{\dagger}$	(0.12)	$0.27^{\dagger}$	(0.12)	$0.31^{\dagger}$	(0.14)
% Avg. impact	$10.35^{\S}$	(1.43)	$3.30^{\dagger}$	(1.39)	$3.51^{\dagger}$	(1.40)	2.32	(1.67)
% Max. impact	$10.67^{\S}$	(1.48)	$3.34^{\dagger}$	(1.44)	$3.60^{\dagger}$	(1.46)	2.29	(1.73)

Robust standard errors in parentheses.
*,† ,§Significant at 10%, 5% and 1% levels, respectively.

Table 5. Regression estimates of causal effect of the contract
By research areas.

Area	Outcome variable	Unconditional ATE		Conditional Restricted ATE		Conditional Unrestricted ATE		Conditional Unrestricted ATT	
Biol.	# papers	0.04	(0.40)	−0.57	(0.48)	−0.51	(0.45)	−1.02*	(0.60)
	Avg. impact	0.69§	(0.18)	0.21	(0.14)	0.21	(0.15)	0.19	(0.23)
	Max. impact	0.96§	(0.26)	0.41*	(0.23)	0.43*	(0.23)	0.37	(0.27)
	% Avg. impact	7.92§	(2.51)	2.70	(2.41)	3.05	(2.51)	−0.41	(2.70)
	% Max. impact	7.50§	(2.57)	2.24	(2.48)	2.61	(2.58)	−1.09	(2.79)
Econ.	# papers	1.01	(0.65)	0.31	(0.85)	0.63	(0.62)	0.21	(0.60)
	Avg. impact	0.12	(0.20)	−0.13	(0.25)	−0.05	(0.21)	−0.24	(0.25)
	Max. impact	0.18	(0.28)	−0.12	(0.36)	−0.14	(0.33)	−0.70	(0.51)
	% Avg. impact	17.88†	(7.50)	13.50	(8.56)	10.20	(9.88)	6.07	(13.23)
	% Max. impact	24.56§	(8.16)	19.42†	(9.25)	15.07	(11.01)	11.78	(16.16)
Chem.	# papers	1.19†	(0.50)	0.49	(0.55)	0.50	(0.56)	0.13	(0.68)
	Avg. impact	0.54§	(0.15)	0.27*	(0.15)	0.32†	(0.14)	0.13	(0.15)
	Max. impact	0.83§	(0.21)	0.50†	(0.22)	0.57§	(0.21)	0.33	(0.22)
	% Avg. impact	13.55§	(3.44)	7.07†	(3.34)	7.68†	(3.32)	3.81	(3.47)
	% Max. impact	13.70§	(3.57)	6.71*	(3.44)	7.33†	(3.44)	3.47	(3.60)
Eng.	# papers	0.46†	(0.23)	0.58†	(0.26)	0.42	(0.29)	0.11	(0.34)
	Avg. impact	0.20§	(0.06)	0.17†	(0.07)	0.14*	(0.08)	0.05	(0.09)
	Max. impact	0.26§	(0.08)	0.25§	(0.09)	0.21†	(0.10)	0.11	(0.12)
	% Avg. impact	8.97§	(3.22)	7.28*	(3.78)	5.06	(4.06)	−0.81	(4.98)
	% Max. impact	9.71§	(3.43)	8.03†	(4.05)	5.41	(4.36)	−1.19	(5.43)
Medic.	# papers	2.38†	(1.13)	1.76	(1.15)	2.93	(1.85)	1.77	(3.38)
	Avg. impact	0.61*	(0.32)	0.09	(0.25)	−0.16	(0.21)	0.18	(0.26)
	Max. impact	0.71	(0.48)	−0.19	(0.41)	−0.54	(0.34)	−0.12	(0.43)
	% Avg. impact	5.75	(4.35)	−1.67	(3.65)	−1.09	(4.04)	−1.29	(4.16)
	% Max. impact	5.95	(4.49)	−1.72	(3.83)	−1.90	(4.07)	−1.56	(4.36)
Phys.	# papers	1.29*	(0.68)	0.90	(0.63)	0.64	(0.50)	1.21	(0.77)
	Avg. impact	1.38§	(0.49)	0.91*	(0.49)	0.60	(0.42)	1.25§	(0.48)
	Max. impact	1.69§	(0.60)	0.92*	(0.55)	0.58	(0.47)	1.29†	(0.55)
	% Avg. impact	15.75§	(4.74)	6.66	(5.17)	4.03	(4.23)	9.17*	(5.20)
	% Max. impact	16.71§	(4.87)	7.57	(5.33)	4.88	(4.42)	10.07*	(5.34)
Hum.	# papers	0.11	(0.22)	−0.02	(0.22)	−0.04	(0.28)	−0.57	(0.35)
	Avg. impact	0.20	(0.14)	0.01	(0.08)	0.01	(0.09)	0.04	(0.14)
	Max. impact	0.33	(0.22)	0.17	(0.17)	0.05	(0.12)	0.25	(0.20)
	% Avg. impact	1.28	(3.29)	−3.15	(2.87)	−3.94	(3.00)	−9.82†	(1.95)
	% Max. impact	1.51	(3.37)	−3.14	(2.97)	−3.97	(3.11)	−10.60†	(5.20)

Robust standard errors in parentheses.

*,† ,§Significant at 10%, 5% and 1% levels, respectively.

Table 6. Matching estimates of causal effect of the contract
Full sample.

Outcome variable	Matching method						Obs.	
	Kernel		NN		Stratification		n_1	n_0
# papers	0.38	(0.26)	0.54	(0.48)	0.13	(0.31)	2147	929
Avg. impact	$0.45^{\S}$	(0.10)	$0.37^{\S}$	(0.13)	$0.33^{\S}$	(0.11)		
Max. impact	$0.58^{\S}$	(0.14)	$0.52^{\S}$	(0.19)	$0.42^{\S}$	(0.15)		
% Avg. impact	$6.43^{\S}$	(1.43)	$5.68^{\S}$	(2.14)	$4.50^{\S}$	(1.47)		
% Max. impact	$6.53^{\S}$	(1.46)	$5.87^{\S}$	(2.08)	$4.53^{\S}$	(1.56)		

Bootstrap standard errors, using 500 replications, in parentheses.
*,† ,§Significant at 10%, 5% and 1% levels, respectively.
n_1, n_0 are the obervations available for treated and controls, respectively.

Table 7. Matching estimates of causal effect of the contract
By research areas.

Area	Outcome variable	Matching method						Obs.	
		Kernel		NN		Stratification		n_1	n_0
Biol.	# papers	−0.56	(0.49)	0.02	(0.99)	−0.61	(0.55)	399	748
	Avg. impact	$0.45^{\S}$	(0.17)	$0.46^{\dagger}$	(0.23)	$0.38^{\dagger}$	(0.19)		
	Max. impact	0.71	(0.28)	0.31	(0.36)	0.63	(0.28)		
	% Avg. impact	4.06^{*}	(2.42)	3.18	(3.96)	4.51	(2.84)		
	% Max. impact	3.53	(2.41)	2.88	(3.99)	4.07	(2.78)		
Econ.	# papers	−1.15	(1.48)	−1.29	(1.96)	0.82	(1.02)	21	38
	Avg. impact	−0.49	(0.51)	−0.19	(0.62)	0.01	(0.31)		
	Max. impact	−0.46	(0.71)	−0.07	(1.12)	0.08	(0.40)		
	% Avg. impact	8.88	(10.9)	25.57	(15.6)	14.87	(9.31)		
	% Max. impact	15.20	(12.42)	30.67^{*}	(16.1)	$21.31^{\dagger}$	(10.0)		
Chem.	# papers	0.65	(0.58)	0.27	(1.08)	0.68	(0.56)	185	281
	Avg. impact	0.34	(0.16)	0.36	(0.23)	0.30^{*}	(0.16)		
	Max. impact	$0.56^{\S}$	(0.21)	0.55^{*}	(0.30)	$0.53^{\dagger}$	(0.23)		
	% Avg. impact	$8.80^{\S}$	(3.36)	$10.23^{\dagger}$	(5.01)	$7.81^{\dagger}$	(3.28)		
	% Max. impact	$8.57^{\S}$	(3.27)	9.99^{*}	(5.12)	$7.59^{\dagger}$	(3.29)		
Eng.	# papers	$0.65^{\dagger}$	(0.28)	0.58^{*}	(0.34)	0.49^{*}	(0.28)	178	185
	Avg. impact	$0.21^{\dagger}$	(0.10)	$0.29^{\dagger}$	(0.13)	0.13	(0.08)		
	Max. impact	$0.29^{\dagger}$	(0.12)	$0.38^{\dagger}$	(0.16)	0.20^{*}	(0.11)		
	% Avg. impact	$9.17^{\dagger}$	(4.54)	$13.48^{\dagger}$	(6.30)	5.15	(4.22)		
	% Max. impact	$10.06^{\dagger}$	(5.02)	$14.55^{\dagger}$	(6.73)	5.73	(4.49)		
Medic.	# papers	1.94	(1.25)	1.49	(1.56)	2.02^{*}	(1.20)	117	271
	Avg. impact	0.28	(0.28)	−0.003	(0.45)	0.24	(0.34)		
	Max. impact	−0.02	(0.44)	−0.89	(0.81)	−0.04	(0.55)		
	% Avg. impact	0.80	(3.90)	−2.79	(6.81)	0.26	(4.67)		
	% Max. impact	0.69	(3.93)	−2.63	(6.96)	0.17	(4.91)		
Phys.	# papers	1.17	(0.75)	1.55^{*}	(0.82)	1.08	(0.78)	93	176
	Avg. impact	$1.25^{\dagger}$	(0.53)	$1.34^{\dagger}$	(0.59)	0.96	(0.48)		
	Max. impact	$1.29^{\dagger}$	(0.58)	$1.46^{\dagger}$	(0.70)	1.01^{*}	(0.52)		
	% Avg. impact	9.19^{*}	(5.19)	12.36	(7.94)	7.73	(5.22)		
	% Max. impact	10.05^{*}	(5.56)	13.34^{*}	(7.75)	8.97^{*}	(5.10)		
Hum.	# papers	0.07	(0.24)	0.28	(0.36)	0.01	(0.23)	80	355
	Avg. impact	0.18	(0.15)	0.28^{*}	(0.17)	0.07	(0.12)		
	Max. impact	0.30	(0.21)	0.44^{*}	(0.25)	0.09	(0.14)		
	% Avg. impact	0.12	(3.30)	4.74	(5.23)	−2.06	(3.01)		
	% Max. impact	0.17	(3.51)	5.26	(5.57)	−2.06	(3.05)		

See Notes to Table 6.

Table A1
Research areas (ANEP)

Physics and Space Sciences
Earth Sciences
Materials Science and Technology
Chemistry
Chemical Technology
Plant and Animal Biology. Ecology
Agriculture
Livestock and Fishery
Food Science and Technology
Molecular and Cell Biology and Genetics
Physiology and Pharmacology
Medicine
Mechanical, Ship and Aeronautical Engineering
Electrical and Electronic Eng. and Robotics
Civil Engineering and Architecture
Mathematics
Computer Sciences
Information and Communication Technologies
Economics
Law
Social Sciences
Psychology and Education Sciences
Philology and Philosophy
History and Art

Table A2
Research areas (Publish or Perish)

Biology, Life Sciences, Environmental Science
Business, Administration, Finance, Economics
Chemistry and Materials Science
Engineering, Computer Science, Mathematics
Medicine, Pharmacology, Veterinary Science
Physics, Astronomy, Planetary Science
Social Sciences, Arts, Humanities