The Impact of French Cancer Research: A New Approach[1]

Gérard Chevalier, Institut National du Cancer

Christine Chomienne, Institut National du Cancer

Nicolas Guetta Jeanrenaud, New York University

Julia Lane, New York University

[1] This paper draws heavily on work with many colleagues, particularly Valerie Thibaudeau and Ghislaine Filliatreau. We have benefited a great deal from the active engagement, empirical contributions, and support of Christian Herzog, of Dimensions, and Laure Haak of ORCID.

Abstract

Much attention has been paid to estimating the impact of scientific investment. However, the effort has been largely manual and hence expensive, burdensome, and error-prone. In addition, the focus has been largely mechanical in nature—drawing a direct line between funding and outputs—rather than focusing on the scientists that do the work.

Although the results should be treated as illustrative, the approach we test here can be used to examine the impact of research funding on scientific output in terms of publications, citations, collaborations and international activity, controlling for both observed and unobserved factors. We argue that full engagement of scientific funders with the scientific community to expand data capacity and evaluation tools would be a fruitful approach to enable a more scientific assessment of scientific investments.

1. Introduction

There is important interest in evaluating the impact of investments in science (*1*, *2*). Part of this is due to the need to justify the relatively high levels of funding, which can be up to 3% of a country's income; part is due to the recognition that technological change and ultimately economic growth relies on investments in research and development and that it is essential to allocate resources as wisely as possible (*3*). However, the empirical evidence has hitherto largely relied on "craft activity" (*4*) and manual reporting due to the lack of an automated systematic data infrastructure for evaluation (*5*). The result has been expensive and too often unconvincing (*6*, *7*). A major reason is that legacy evaluation approaches have been focused on capturing information on documents, rather than on the scientists who were funded. As such, it is not possible to either construct comparison groups or control for the many unobserved factors that contribute to scientific productivity.

This paper describes how modern data-driven approaches and statistics can be used to improve the evidence basis for research evaluation. It uses the concrete example of an evaluation of an agency funding cancer research in France—the Institut National du Cancer (INCa)—to illustrate the method. The context is similar to that of many other science agencies. In coordination with other public institutions and charities[2], INCa allocates around 100 million euros per year to research projects through a standard mechanism of calls for proposals[3]. As a relatively young institution, established in 2004, INCa set up a number of procedures and tools to manage its grants. However, like many other science agencies, the data infrastructure around its grant-making was solely legal and administrative. When asked to evaluate the impact of INCa investments, senior management found that its in-house ability to do so was severely limited.

An advantage of being a relatively young organization was that INCa management could examine modern approaches to examining impact. In addition, cancer research is a particularly appealing initial case study, because there are international common standard taxonomies for cancer and so comparisons can be made to other cancer funding agencies. Thus, in 2012, INCa launched a pilot

[2] Direction générale de l'offre de soins (Ministère de la santé), Alliance pour les sciences de la vie et de la santé (ITMO Cancer), Ligue contre le cancer, Fondation ARC.

[3] As a word of caution, comparing amounts from one country to another is a tricky exercise: in France, the amounts allocated to research projects by INCa and similar entities do not cover the researchers' salaries and other costs which are usually required by other countries' research funding systems, for example in the USA.

project named HELIOS (Health Investments Observatory) to make use of its administrative data and link it with available publication and patent databases. The pilot confirmed the feasibility of the approach and contributed to identify the building blocks of an integrated system which could be used to assess the impact of INCa funding in the long term (i.e., long after the completion of the projects). The success of this pilot project was acknowledged when the 2014–2019 National Cancer Plan mandated INCa to "develop shared tools for the evaluation of research projects in oncology"[4]. The main funders of scientific and clinical research in France were therefore invited to collaborate and responded with great interest.

In this study we describe how the approach was implemented, and more importantly, how the approach is replicable, low-cost, and can be scaled to any other research funder. We find that the new data infrastructure has the scientific foundations necessary to support high-quality impact evaluations, particularly in the case of cancer research. Although the results should be treated as illustrative, the approach can be seen as the basis for the scientific analysis of the impact of research funding on scientific output in terms of publications, citations, collaborations and international activity, controlling for both observed and unobserved factors. We find that full engagement of scientific funders with the research community to expand data capacity and evaluation tools would be a fruitful approach to enable a more scientific assessment of scientific investments.

2. Background

The INCa context

In considering how to evaluate the impact of their investments, the French agencies agreed to two core principles. First and foremost, the public entities as well as the private charities who actively participated in this process[5] agreed that building such tools should not be done at the expense of

[4] Action 17.13: Développer des Outils Partagés d'Evaluation des Projets de la Recherche en Cancérologie (https://www.e-cancer.fr/Plan-cancer/Plan-cancer-2014-2019-priorites-et-objectifs/Plan-cancer-2014-2019-de-quoi-s-agit-il/Les-17-objectifs-du-Plan/Objectif-17-Adapter-les-modes-de-financement-aux-defis-de-la-cancerologie)

[5] Agence Nationale de la Recherche; Agence de Biomédecine; Agence nationale de Sécurité du Médicament et des Produits de Santé; Agence Nationale de Recherches sur le Sida et les Hépatites Virales; Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail; Alliance Nationale pour les Sciences de la Vie et de la Santé; Direction Générale de la Recherche et de l'Innovation; Direction Générale de l'Offre de Soins; Direction Générale de la Santé; Haut Conseil de l'Evaluation de la Recherche et de l'Enseignement Supérieur; Association Française contre les Myopathies; Ligue contre le Cancer; Fondation ARC pour la Recherche sur le Cancer; Fondation de France; Fédération Française de Cardiologie; Fondation pour la Recherche Médicale; Institut de Recherche pour le Développement; Fédération Française de Cardiologie; France Alzheimer et Maladies Apparentées.

the researchers and research institutions, who should be left unburdened to concentrate on their scientific activities. This stands in sharp contrast to the UK Research Excellence Framework, which has been estimated to cost UK institutions almost £250 million, and about £4,000 per submitted researcher.

The second principle addresses the needs of the research funders in charge of the assessment. Traditionally, they rely on unstructured reports written by researchers at the end of their funded projects and, for special purposes, on additional reports requested from the researchers after a longer time period. To extract relevant information from such reports is a painstaking exercise and subject to many biases. Well-structured databases and robust assessment methodologies should eliminate these flaws. The 2012 HELIOS pilot project showed that the construction of automated databases required the definition of standards, particularly consistent identifiers and ways of classifying research across agencies. Once integrated in grants management systems, such conventions facilitate the data extraction and its linkage. French funders agreed to establish, at the national level, recommendations on such standards and have formed working groups to develop white papers supporting the recommendations.

Data

Legacy systems have quite limited information on *what* research is funded, and certainly do not have information on what other funders are doing. It has historically been a herculean task to manually pull such data together from multiple sources and standardize the information. However, new data sources have become available, such as Dimensions[6], a Digital Science database tool aggregating publications, citations, grants, clinical trial patents and policy papers. Using advanced techniques of Natural Language Processing and Machine Learning, the database connects research metadata such as researcher profiles, grants, and publications of all types. Dimensions now comprises structured data relating to more than 4 million funded projects, 98 million scientific publications, and more than 1 billion citations. The Dimensions team also linked in the INCa data to the Dimensions database, and used the database to capture information on other cancer funders and researchers.

---

[6] https://www.digital-science.com/products/dimensions/ and https://app.dimensions.ai for direct access. Dimensions applies standard preprocessing and normalization techniques to disambiguate funders and researchers.

Identifying research topics is a major challenge. Among the many features included in the Dimensions database is the grant's or publication's topic, as defined by the Research, Condition, and Disease Categorization (or RCDC) categorization system[7]. Initially developed by NIH, the RCDC process used Machine Learning classification to create 233 carefully crafted topics, or categories. Over the past 10 years, RCDC categories were coded to all NIH grants based on the grant content. The Digital Science team, which was involved in the creation of RCDC, has more recently developed a Machine Learning approach to automatic classification of non-NIH grants, and included it on the Dimensions platform. Using the coded grants as a training set, RCDC categories were assigned to all publications and grants on the Dimensions database. In addition, Dimensions provided more detailed cancer specific codes, called Common Scientific Outline (CSO) codes developed by the International Cancer Research Partnership[8].

Another major challenge is to trace the research activity of scientists both before and after the award (i.e., the initial results of funding). In order to facilitate this, we worked closely with the ORCID[9] organization. ORCID is an established researcher identifier registry used by over 6 million researchers. ORCID enables individuals to register for a unique identifier, and connect it with their activities and affiliations in common research workflows such as grant application, publication submission, peer review, and dataset deposit. Researchers control their record and may share their information publicly. Many research funders are starting to adopt the use of ORCID, including INCa and the US National Institutes of Health, and some require the use of ORCID in grant application workflows, including the Wellcome Trust and the UK National Institutes of Health Research[10].

Impact Measurement

Our review of the literature identified three areas key to measuring impact in the context of science. The first is conceptual—focusing on people, rather than documents. The second is measurement—building better ways to capture data. The third is statistical—developing comparison groups and adjusting for selection bias.

---

[7] https://report.nih.gov/rcdc/
[8] https://www.icrpartnership.org/
[9] https://orcid.org/
[10] https://orcid.org/organizations/funders/policies

The conceptual framework has evolved over the past decade to focus on people, rather than documents (*8*, *9*). The mechanical reliance on bibliometrics in general and scientiometrics in particular has been harshly criticized as slow, narrow, secretive and irreproducible, open to gaming and inappropriate in scope (*10*). At a conceptual level, the more recent literature recognizes that people and networks are the drivers of innovation (*11*, *12*), and thorough analysis increasingly points to the importance of intangible flows of knowledge, such as contacts at conferences, business networking, and student flows from the bench to the workplace (*13*).

Measurement issues are a major challenge since there are poor current measures of inputs (all the individuals who are funded, the funding levels, the structure and duration of funding), of the units of analysis (networks, project teams, collaborations), and of innovation measures (patents, publications, new products and processes) (*13–17*).

At a statistical level, the evaluation of a scientific hypothesis—like a funding intervention—requires comparing treatment and control groups. Cancer funding, like other scientific investments, is typically based on a set of selection criteria, so it is important to adjust for non-random participation and identify an appropriate comparison group (or groups). The standard impact evaluation framework is to determine what is the impact (Δ) or causal effect of a program (P) on an outcome of interest (Y):

$$\Delta = (Y \mid P=1) - (Y|P=0)$$

In other words, the causal impact (Δ) of a program (P) on an outcome (Y) is the difference between the outcome (Y) with the program (Y=1) and the same outcome (Y) without the program (when P=0) (*1*). The approach is to describe outputs, confounding factors, and develop a counterfactual. In the case of research funding, a theory of change would be that research funding works to attract scientists, or teams of scientists, to study the topic of interest to the funder. The funding pays for both people's time and for research inputs, like equipment, materials and physical or scientific infrastructure. The result of combining people and other inputs is the creation of new scientific ideas, together with their dissemination and subsequent adoption in a variety of arenas—other scientific fields, business activity, clinical activities or policy.

Figure 1 provides an illustrative overview of the conceptual framework we used: research funding pays for the Principal Investigator (and their institution) to pay for people's time and scientific inputs, which are then combined to create outputs.

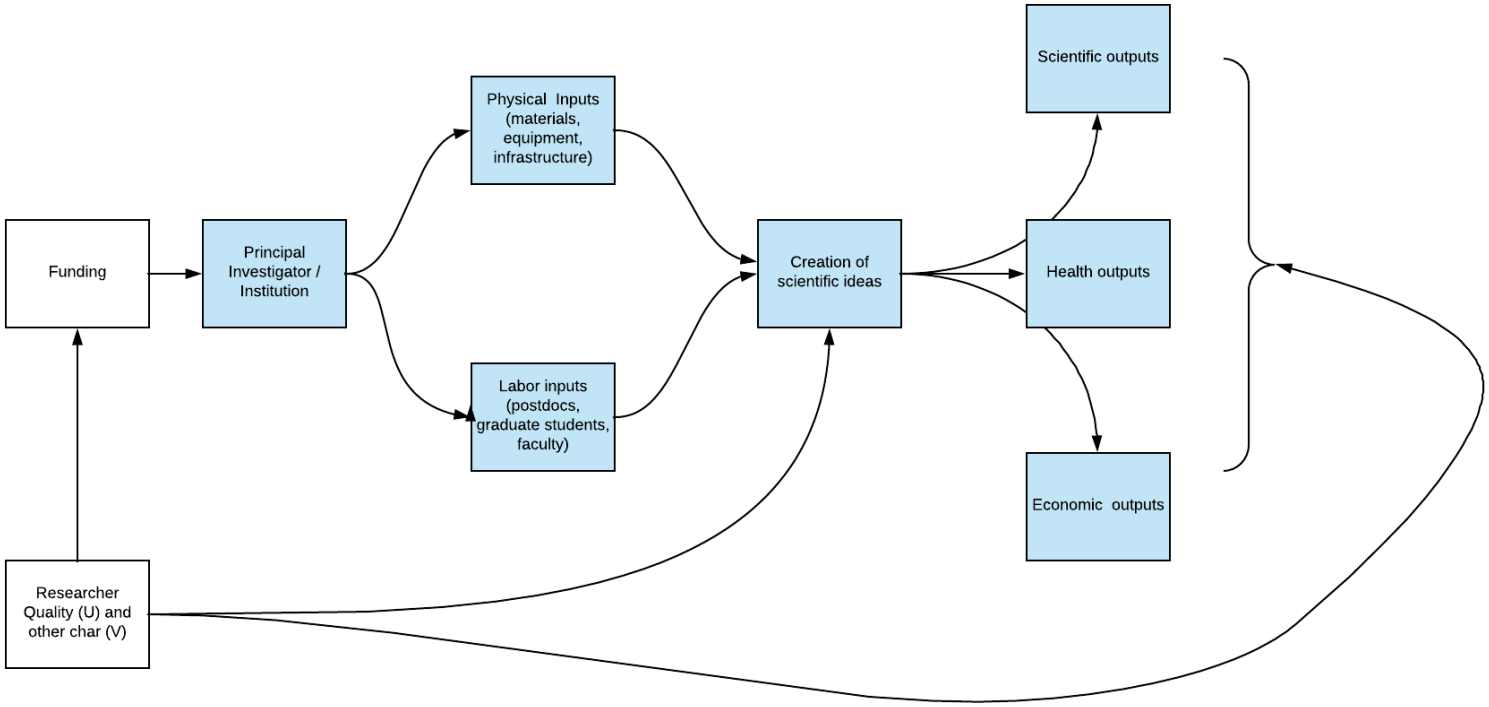

Figure 1: Conceptual Framework

Of course, this diagram is overly simplistic. Science is nonlinear and complex, with long and often complicated causal chains—just like any other human activity, like education, criminal activity, or employment. Indeed, there is a considerable literature on the statistical issues associated with estimating impact in all of these areas, on which we draw (*18*, *19*). INCa's funding, like that of most other funding agencies, is typically predicated on a peer review process that funds the "best" research, which creates a fundamental evaluation problem due to selection bias (*23*). Writing down the process in terms of the framework in Figure 1, scientists are awarded funding based on panel review, X, and other individual characteristics, V. To investigate the effect of the funding on science, Y, one may be concerned about the potential confounding effect of the scientist's quality, U, which may not be precisely measured by X[11]. Positive selection (i.e., if higher quality

[11] This description is directly taken from Abadie and Cattaneo (*18*).

researchers are more likely to be selected for funding) is likely to upwardly bias estimates of the effect of science funding.

One popular approach to estimating impact is to construct comparison groups consisting of scientists as similar as possible to the funded group, but who do not receive funding. This approach requires having reasonable measures of the confounding covariates, U and V. This approach has been extensively applied since its introduction by Rosenbaum and Rubin (*27*, *28*), and essentially groups similar individuals based on the propensity to be treated (see Smith and Todd (*35*) and Caliendo and Kopeinig (*36*) for good discussions)[12]. The National Institutes of Health has also used it in different contexts (*21*). There are several canonical approaches. One is to find a "nearest neighbor" based on similar covariates. Another is to create "propensity weights" which reflect the conditional probability of receiving the treatment given measurable covariate values. Others combine variations of these approaches.

Sometimes, however, the confounding covariates are not measurable. In this case, the canonical approach is to apply "difference in difference" estimators where the confounding factors are assumed to be individually invariant and hence the differences in levels between groups are eliminated if regressions are estimated in terms of changes or controlling for fixed effects rather than levels. This permits the identification of change between two time periods resulting from the intervention. This methodology is particularly useful when it is not possible to observe directly the characteristic of a population (persons, firms, etc.) participating in a program.

A new technique which has been developed that is of great interest in this context is called a "synthetic" control approach. This approach constructs a weighted average of individuals who do not receive funding in the base period, and compares them to individuals who do receive funding. This approach is arguably the most important innovation in the policy evaluation literature in the last 15 years (*35, 36*).

[12] A technical summary is provided in the Supplementary Online Materials.

3. Making comparisons

The core approach used in this paper was to develop data that described what was funded, who was funded, and the results—relative to a comparison group. We work with data from five different agencies in four countries: the United States National Cancer Institute, Cancer Research UK, Wellcome Trust, the Australian National Health and Medical Research Council, and cancer research funded programs by the French National Cancer Institute (INCa), the French National Alliance of Life Sciences and Health (AVIESAN) through the Institut National de la Santé et de la Recherche Médicale (INSERM) and the Ministry of Health through its Direction Générale de l'Offre de Soins (DGOS) between 2007 and 2012. Two of the agencies are cancer-specific (Cancer Research UK and National Cancer Institute): for those, all awards are considered. Wellcome Trust and the National Health and Medical Research Council however are general funding bodies for all medical research: in order to restrict to cancer-related grants, we used a machine learning classification process based on a system developed for the U.S. National Institutes of Health (the RCDC classification, see Section 2.).

Overall, our analysis accounts for 9,922 grants, awarded between 2007 and 2012 to 12,083 unique researchers (subject to the disambiguation algorithm); 859 from CRUK, 914 from INCa, 8,521 from NCI, 1,543 from NHMRC and 246 from Wellcome Trust. We exclude exceptional cases of researchers funded by several of the five agencies and consider only career-first grants, so all researchers of the data unique and associated to only one grant (subject to the reporting biases noted above).

A simple comparison of one output measure—the average annual publications of researchers funded by the different funding agencies—shows remarkable differences both in the relative productivity of researchers across funding agencies and in the net change in productivity subsequent to receiving funding. An examination of Figure 2 shows that INCa funders (comprising INCa, ITMO Cancer and DGOS) are both more productive by this measure than other funders (an average output of 5.7 publications a year, compared with as few as 3.3 for Wellcome Trust researchers and 3.0 for NCI researchers) and that the funding increased researchers' productivity more than others funding (2.5 publications per year compared with roughly 1.5 to 2 for researchers of other funding agencies).

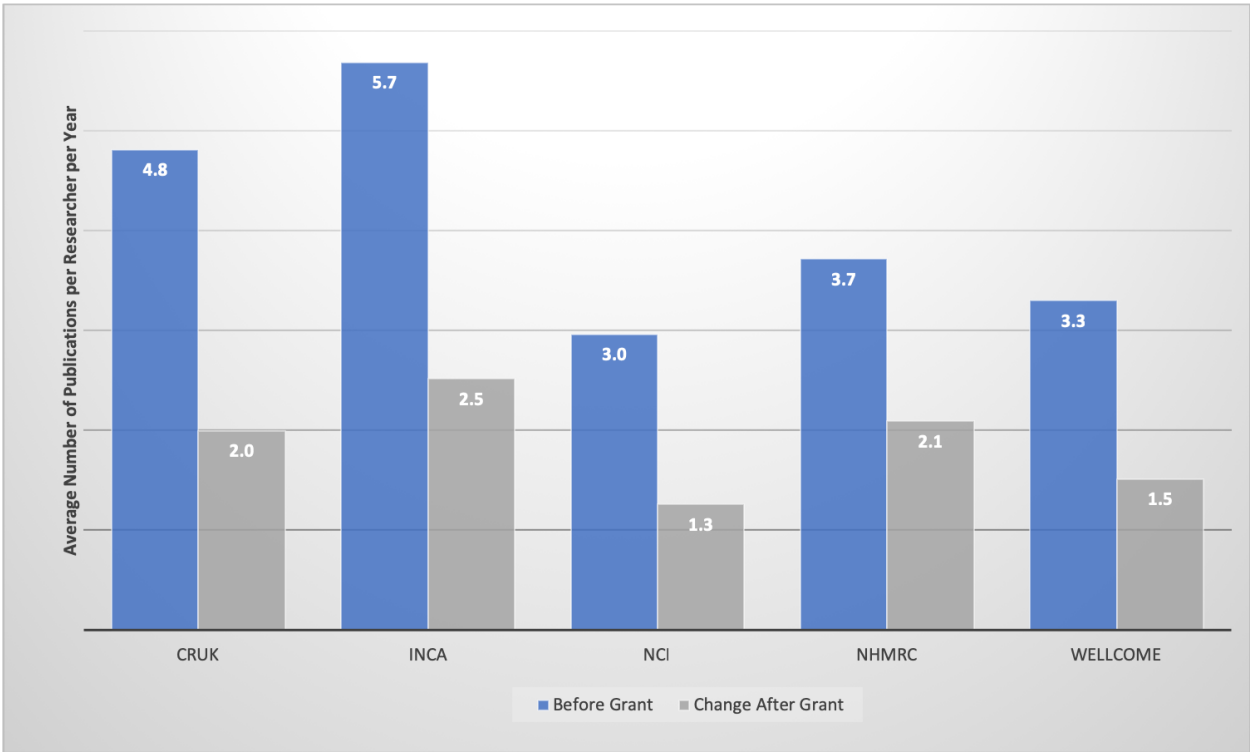

Figure 2: Annual Researcher Publications by Funding Agency

Of course, these results are not directly comparable because different cancer funding agencies have both different selection criteria, different research foci and different funding levels—each of which can result in different levels of publications. Fortunately, statistical techniques, such as propensity weights can be applied to normalize both the input and output variables to make the comparisons more reasonable.

In order to construct reasonably comparable groups, we need to capture as much information about researchers, their funding, and their research fields as possible. We use the data structure of the analytical sample to include information about scientific activity for each funded researcher in the five years before, during, and after their first grant, yielding 11 observations per researcher for each of the funding agencies[13]. We include information about the researcher's career age, gender and affiliation. The propensity weights are created using number of publications in the first year, the RCDC and CSO codes of these publication, and the career age.

[13] The full discussion of the sample and variable construction, the construction of the propensity weights and the full results are presented in the appendices.

In what follows, we focus on examining the impact of cancer funding in each of the comparator funding agencies for five output measures of interest: publications, citations, publications in the same field as the grant, collaborators, and international collaborations. In practical terms, this approach can enable each funder to compare the performance of their researchers to those of any of the other funders—or a weighted combination if so desired.

Figure 3 graphically demonstrates the results of the adjustment using propensity weights. As can be seen, the dramatic differences in productivity of researchers across funders is substantially attenuated—and, indeed, the point differences between CRUK, NCI, and NHMRC are not statistically significant. Similarly, the changes in productivity subsequent to funding are much smaller—apart from Wellcome Trust, all agencies show an increase of over 1.9, similar to INCa's 2.5 publications per year increase.

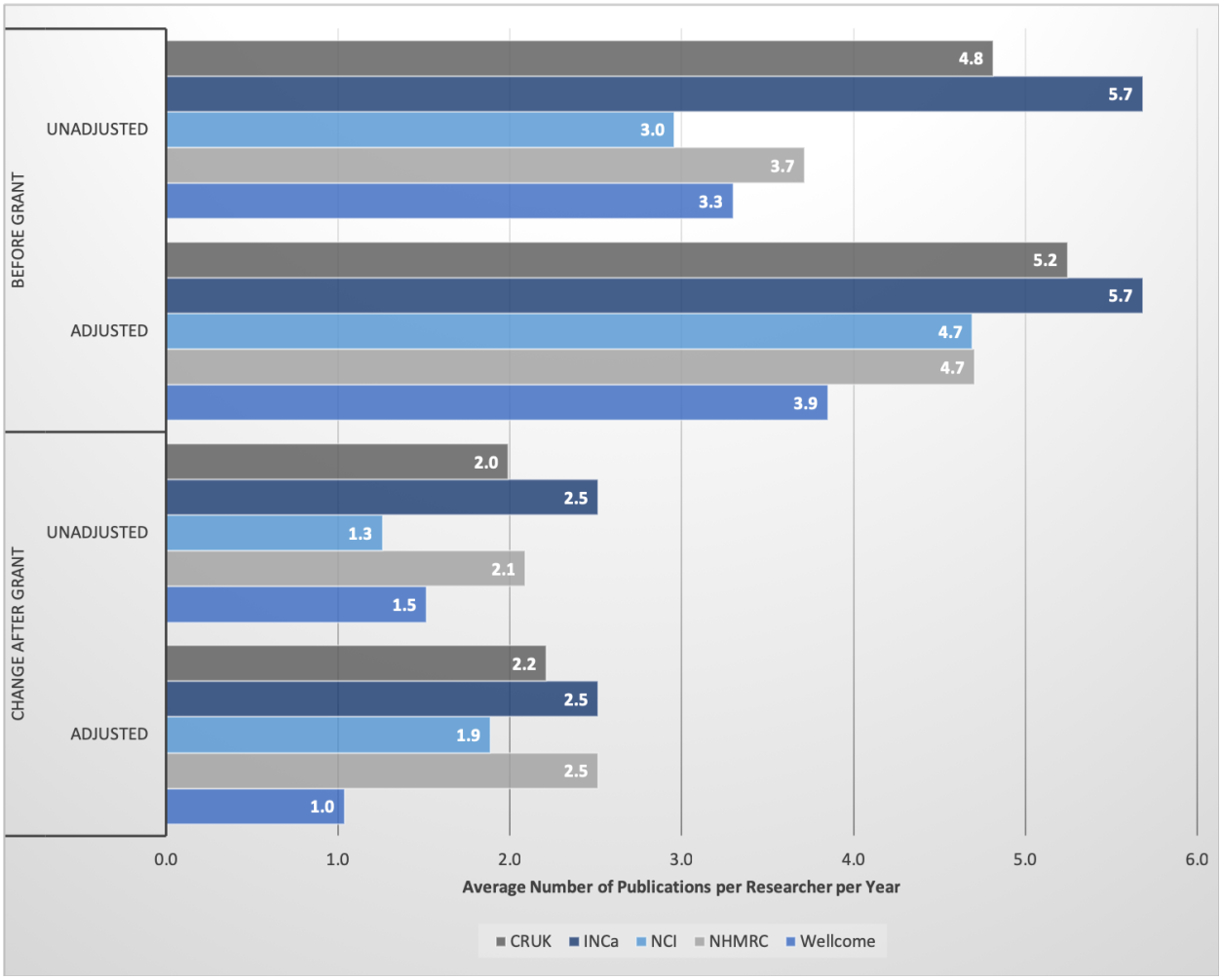

Figure 3: Annual Researcher Publications by Funding Agency

The regression results presented in Appendix F of the Supplementary Online Material differ between specifications because there are controls for demographic characteristics and for research type. Interestingly, one potentially important finding is that there is a negative coefficient associated with being a female researcher in almost all the model specifications. Although the focus of this paper is not to discuss gender differences, it is worth noting that this result is consistent with a burgeoning literature documenting substantive differences in the relative positions of women and men in scientific networks, collaborations, and organizations. Whittington and Smith Doerr study a sample of academic and industrial life scientists and show that women are less likely to patent than men even after controlling for education and career interruptions. However, in organizational settings like biotechnology firms that have flatter organizational structures than industry and academia, women are more likely to become patent holding inventors (*22*). There is a literature that describes pathways whereby women might be placed in different positions than men, since there are specific channels that link individuals (*23*). The consequence of different positioning in networks is likely to be that there is a longer search process to find a key expert, and it is consequently more difficult to identify and develop true innovations (*24*).

As the number of publications is a relatively crude output measure, it is important to know whether or not the science in the publication is cited. With the exception of NHMRC, once the analysis is adjusted, there is no difference across funders in terms of citations. And, while there is a positive increase in citations subsequent to funding, the only group that is different from the others is NCI. Interestingly, though the straight difference in differences would suggest that research funding increases the visibility (i.e., citation rate) of a researcher's work, using a more sophisticated set of controls shows that there is no significant effect on the citation rate, whether for funders in general or for INCa funding in particular.

Other outcome measures, such as the number of collaborators, the extent of international collaborations and the closeness of publication topic to the grant topic, were examined; the full results are presented in the Supplementary Online Materials. One of the most interesting data, made possible by text analysis, was whether researchers funded by different agencies published articles in the same RCDC Code topic as the grant. The benchmark researcher, prior to getting a grant, published an average of just fewer than 3 articles in the same field as the grant. However, the average INCa researcher had almost 1 fewer; the NCI had almost 1.3 more. The number of publications in the research field increased by about 1 per year after the grant; the only funders

that saw significant increases in field related publications that were robust across multiple specifications were NCI and NHMRC—which may well be due to the larger sample size for these funders.

The differences in the outcome measures across funders—as well as the differences in the characteristics of the researchers prior to funding—highlight the important issues associated with doing an evaluation of this type. First, it is clear that during the period analyzed (2007–2012) the different funders specialized in different areas. Wellcome Trust seems to specialize more in basic research why INCa was more in applied research. Publication patterns may well be very different across these areas, and developing measures to normalize those differences would be an important step to ensure the robustness of the finding. Second, funding is not exogenous. Each funder has a different selection process, and the selection process is likely to be one of the unobservable factors that we highlighted in Figure 1. Funders could share data on both those who are funded, and those who are not funded to adjust for such differences—this adjustment is called a regression discontinuity approach (*27*). In addition, there may be systemic differences in the quality of reporting across funders. For example. NCI does not require an ORCID ID, which may result in a lower high-quality information about the NCI funded researchers' publications.

4. What more could be done

The focus of this paper has been to document the potential for a new approach that can be used to systematically describe the results of investment in research. It shows how new data about research funding, researchers, and researchers' scientific activity can be combined to create a new scientific data infrastructure to study the activities of scientists—a science of science. It applies statistical techniques for program evaluation to examine the relative impact of French funding for cancer research. The results presented in this study should not be used as results per se but as illustrations of an approach that highlights the relevance of controlling for both observed and unobserved factors when examining the impact of research funding on scientific output in terms of publications, citations, collaborations and international activity. Much more can be done. With thorough statistical methods at hand, and collected data, the scientific community could develop more outcome measures, such as student placements, data and code sharing activities, interdisciplinary and related research.

There is clearly momentum to move in this direction. In the United States, the expansion and growth of the STAR METRICS/UMETRICS approach has been instantiated in the establishment of the Institute for Research on Innovation and Science (*28*). That work has been coupled with the Innovation Measurement Initiative at the US Census Bureau (*29*) and has led to deeper understanding of how research activity stimulates economic and scientific innovation (*30–32*). In Australia, a new recommendation from Australian Parliament calls for the use of ORCID and streamlined reporting to enable evaluation (*33*). There is also a groundswell of support for the use of identifiers and standards by both the EC and in the French Open Science policy (*34*). We hope that the work reported here will also stimulate funding agencies to adopt similar approaches, and engage with the scientific community, to facilitate a much more scientifically oriented, reproducible, and evidence-based approach to the assessment of scientific investments.

References

1. J. Marburger, Wanted: better benchmarks. *Science (1087)*. **308** (2005), in press, doi:10.1126/science.1114801.
2. B. S. Bernanke, Promoting Research and Development: The Goverment's Role. *Issues Sci. Technol.* **27**, 37–41 (2011).
3. P. M. Romer, Endogenous technological change. *J. Polit. Econ.* **98**, S71–S102 (1990).
4. B. R. Martin, The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster? *Res. Eval.* **20**, 247–254 (2011).
5. National Science and Technology Council, "The Science of Science Policy: A Federal Research Roadmap" (National Science and Technology Council, Science of Science Policy Interagency Task Group, Washington, D.C., 2008).
6. C. MacIlwain, Science economics: What science is really worth. *Nature*. **465**, 682–684 (2010).
7. T. Penfield, M. J. Baker, R. Scoble, M. C. Wykes, Assessment, evaluations, and definitions of research impact: A review. *Res. Eval.* (2013) (available at http://rev.oxfordjournals.org/content/early/2013/10/08/reseval.rvt021.abstract).
8. J. I. Lane, A call to action to build research data infrastructure. *Nat. Hum. Behav.* **1**, 75 (2017).
9. J. I. Lane, Watch the Players, not the Scoreboard. *Nature* (2017).
10. J. Priem, B. Hemminger, 2.0: Toward new metrics of scholarly impact on the social Web. *First Monday*. **15** (2010).
11. K. B. Whittington, J. Owen-Smith, W. W. Powell, Networks, propinquity, and innovation in knowledge-intensive industries. *Adm. Sci. Q.* **54**, 90–122 (2009).
12. W. Powell, E. Giannella, in *The Handbook of Innovation* (Elsevier, Amsterdam, 2010).
13. C. Corrado, J. Haskel, C. Jona-Lasinio, Knowledge spillovers, ICT and productivity growth. *Oxf. Bull. Econ. Stat.* (2017).
14. J. Mairesse, P. Mohnen, E. Kremp, The Importance of R&D and Innovation for Productivity: A Reexamination in Light of the French Innovation Survey. *Ann. Econ. Stat. / Ann. d'Économie Stat.* (2005).
15. J. Mairesse, P. Mohnen, in *Handbook of the Economics of Innovation- Vol-II,* B. Hall, N.

Rosenberg, Eds. (Academic Press, 2010).

16. A. Jaffe, B. Jones, *The Changing Frontier: Rethinking Science and Innovation Policy* (University of Chicago Press, 2014; http://www.nber.org/books/jaff13-1).
17. C. Corrado, J. Lane, "Using Cyber-enabled Transaction Data to Study Productivity and Innovation in Organizations," *Global COE Hi-Stat Discussion Paper Series 099* (The Conference Board, 2009).
18. A. Abadie, M. D. Cattaneo, Econometric methods for program evaluation. *Annu. Rev. Econom.* **10**, 465–503 (2018).
19. S. Athey, G. W. Imbens, The state of applied econometrics: Causality and policy evaluation. *J. Econ. Perspect.* **31**, 3–32 (2017).
20. P. Azoulay, Research efficiency: Turn the scientific method on ourselves. *Nature*. **484**, 31–32 (2012).
21. L. L. Haak, J. Schnell, "National Institutes of Health Individual Mentored Career Development Awards Program" (Bethesda, MD, 2011), (available at https://grants.nih.gov/training/K_awards_evaluation_finalReport_20110901.pdf).
22. K. B. Whittington, L. Smith-Doerr, Women inventors in context: Disparities in patenting across academia and industry. *Gend. Soc.* **22**, 194–218 (2008).
23. S. Breschi, F. Lissoni, Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *J. Econ. Geogr.* **9**, 439–468 (2009).
24. J. Singh, M. T. Hansen, J. M. Podolny, The world is not small for everyone: Inequity in searching for knowledge in organizations. *Manage. Sci.* **56**, 1415–1438 (2010).
25. R. Crescenzi, M. Nathan, A. Rodríguez-Pose, Do inventors talk to strangers? On proximity and collaborative knowledge creation. *Res. Policy*. **45**, 177–194 (2016).
26. Direct Science, H. Draux, S. Kundu, Gender Imbalance in Cancer Research Grants (2018), doi:10.6084/m9.figshare.7378001.v1.
27. R. Bronzini, E. Iachini, Are Incentives for R&D Effective? Evidence from a Regression Discontinuity Approach. *Am. Econ. J. Econ. Policy*. **6**, 100–134 (2014).
28. J. Lane, J. Owen-Smith, R. Rosen, B. Weinberg, "New linked data on science investments, the scientific workforce and the economic and scientific results of science" (2014).
29. J. Lane, J. Owen-Smith, J. Staudt, B. A. Weinberg, in *Center for Economic Studies and*

Research Data Centers Research Report: 2017, US Census Bureau, Ed. (Washington DC, 2018).

30. N. Goldschlag, R. S. Jarmin, J. I. Lane, N. Zolas, in *Measuring and Accounting for Innovation in the 21st Century*, C. Corrado, J. Haskel, J. Miranda, D. E. Sichel, Eds. (University of Chicago Press, 2018).
31. N. Zolas *et al.*, Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D. recipients. *Science (80–.).* **350** (2015), doi:10.1126/science.aac5949.
32. J. Owen-Smith, *Research Universities and the Public Good: Discovery for an Uncertain Future* (Stanford University Press, 2018).
33. Australian Parliament, "Australian Government Funding Arrangements for non-NHMRC Research" (Canberra, Australia, 2018), (available at https://www.aph.gov.au/Parliamentary_Business/Committees/House/Employment_Education_and_Training/FundingResearch/Report/section?id=committees/reportrep/024212/26656).
34. J. Bosman, Innovations in Scholarly Communication: Changing Research Workflows (2018), (available at https://twitter.com/i/web/status/981610644612046848).
35. J. A. Smith, P. E. Todd, Reconciling conflicting evidence on the performance of propensity-score matching methods. *American Economic Review.* **91.2**, 112–118 (2001).
36. M. Caliendo, S. Kopeinig, Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys.* **22.1**, 31–72 (2008).