



RESEARCH ARTICLE

A new approach for estimating research impact: An application to French cancer research

G rard Chevalier¹, Christine Chomienne^{1,2}, Nicolas Guetta Jeanrenaud³,
Julia Lane³, and Matthew Ross³

¹Institut National du Cancer

²ITMO Cancer Aviesan

³New York University

an open access  journal



Citation: Chevalier, G., Chomienne, C., Jeanrenaud, N. G., Lane, J., & Ross, M. (2020). A new approach for estimating research impact: An application to French cancer research. *Quantitative Science Studies*, 1(4), 1586–1600. https://doi.org/10.1162/qss_a_00087

DOI:
https://doi.org/10.1162/qss_a_00087

Supporting Information:
https://doi.org/10.1162/qss_a_00087

Received: 8 April 2019
Accepted: 24 August 2020

Corresponding Author:
Julia Lane
julia.lane@nyu.edu

Handling Editor:
Ludo Waltman

Copyright:   2020 G rard Chevalier, Christine Chomienne, Nicolas Guetta Jeanrenaud, Julia Lane, and Matthew Ross. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: HELIOS, research impact, science data infrastructure, science impact

ABSTRACT

Much attention has been paid to estimating the impact of investments in scientific research. Historically, those efforts have been largely ad hoc, burdensome, and error prone. In addition, the focus has been largely mechanical—drawing a direct line between funding and outputs—rather than focusing on the scientists that do the work. Here, we provide an illustrative application of a new approach that examines the impact of research funding on individuals and their scientific output in terms of publications, citations, collaborations, and international activity, controlling for both observed and unobserved factors. We argue that full engagement between scientific funders and the research community is needed if we are to expand the data infrastructure to enable a more scientific assessment of scientific investments.

1. INTRODUCTION

There is great interest in evaluating the impact of investments in science (Bernanke, 2011; Marburger, 2005). 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 rely on investments in research and development and that it is essential to allocate resources as wisely as possible (Romer, 1990). However, the empirical evidence has hitherto largely relied on “craft activity” (Martin, 2011) and manual reporting due to the lack of an automated systematic data infrastructure for evaluation (National Science and Technology Council, 2008). The result has been expensive and too often unconvincing (MacIlwain, 2010; Penfield, Baker, et al., 2014). A major reason is that legacy evaluation approaches have focused on capturing information on documents, rather than on the scientists who received funding. Therefore, it is not possible to either construct comparison groups or control for the many unobserved factors that contribute to scientific productivity at an individual level.

This paper describes a modern data-driven approaches and empirical methodology that can be used to improve evidence-based research evaluation. It provides an illustrative example of the utility of these methods by evaluating an agency funding cancer research in France—the Institut National du Cancer (INCa). The context is similar to that of many other science agencies. In coordination with other public institutions and charities¹, INCa allocates around  100 million per

¹ Direction g n rale de l’offre de soins (Minist re de la sant ), Alliance pour les sciences de la vie et de la sant  (ITMO Cancer), Ligue contre le cancer, Fondation ARC.

year to research projects through a standard mechanism of calls for proposals². 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 capacity was limited.

An advantage of being a relatively young organization was that INCa management could examine modern approaches to evaluating impact. In addition, cancer research is a particularly appealing initial case study, because there are standard international taxonomies for cancer and so comparisons can be made to other cancer funding agencies. Thus, in 2012, INCa launched a pilot 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 that 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.”³ The main funders of scientific and clinical research in France were therefore invited to collaborate and responded with great interest⁴.

In this study, we describe how the individual-centered approach was implemented and, more importantly, how the approach is replicable, low cost, and easily applicable to other research funders. 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. In particular, the data allow for evaluating scientific output by tracing out the links between grants to individuals and their subsequent activity 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

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

² 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 that are usually required by other countries' research funding systems, for example in the United States.

³ Action 17.13: Développer des Outils Partagés d'Évaluation 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>).

⁴ 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'Évaluation 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.

measurement—building better ways to capture data. The third is statistical—developing comparison groups and adjusting for selection bias.

The *conceptual* framework has evolved over the past decade to focus on people, rather than documents (Powell & Giannella, 2010; Whittington, Owen-Smith, & Powell, 2009). Our work builds on important earlier work, which has focused on the importance of individuals and teams. For example, work by Bozeman and coauthors has stressed the importance not only of individual human capital endowments but also researchers' know-how in terms of both their tacit and craft knowledge (Bozeman, Dietz, & Gaughan, 2001) and transdisciplinary collaboration networks (Bozeman & Rogers, 2002). Other work includes the project SIAMPI, which examined the interactions between researchers and users (Spaapen & Van Drooge, 2011), as well as the work done in the project ASIRPA, which used ex post analysis of the networks of interactions to describe how results were achieved (Joly et al., 2015).

Recent research has increasingly emphasized the importance of intangible flows of knowledge, such as contacts at conferences, business networking, and student flows from the bench to the workplace (Corrado, Haskel, & Jona-Lasinio, 2017). However, measurement issues are a major challenge, as 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) (Corrado & Lane, 2009; Corrado et al., 2017; Jaffe & Jones, 2014; Mairesse & Mohnen, 2010; Mairesse, Mohnen, & Kremp, 2005).

The *measurement* should also be automated to reduce cost and increase transparency. Building a data infrastructure should not be done at the expense of 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. This means less reliance 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 period. To extract relevant information from such reports is a painstaking exercise and subject to many biases. The 2012 HELIOS pilot project showed that the construction of automated databases required the definition of standards, particularly consistent identifiers such as ORCID, to trace researchers, and ways of classifying research across agencies, such as the Areas of Research established by the International Cancer Research Partnerships. Once integrated in grant management systems, such conventions facilitate data extraction and linkages. French funders agreed to establish, at the national level, recommendations on such standards and have formed working groups to develop white papers supporting the recommendations⁵.

The third area is developing *comparison* groups. Indeed, there is a considerable literature on the statistical issues associated with estimating impact by constructing plausible comparison groups (Abadie & Cattaneo, 2018; Athey & Imbens, 2017). Science funding is typically predicated on a peer review process that funds the “best” research, which creates a fundamental evaluation problem due to selection bias (Breschi & Lissoni, 2009).

A relatively newly developed technique that is of great interest in this context is called a “synthetic” control. This technique combines the canonical propensity score and difference-in-difference methods. Specifically, the researcher constructs a control group consisting of untreated individuals weighted using inverse propensity scores. Distinct from the canonical

⁵ <https://www.e-cancer.fr/Professionnels-de-la-recherche/Paysage-de-la-recherche/Le-projet-HELIOS>

propensity score method, these individuals need not actually have been eligible for receiving funding from the focal agency. Further, the method relies on difference-in-differences to compare those who receive funding from the focal agency to those who receive funding from a different agency, but after adjusting the control group based on an apples-to-apples weighting of covariates in the pretreatment period. Because the control is limited to those having received funding from other non-French agencies, rather than completely untreated individuals, there is less of a concern about selection on time variant unobservables. Overall, the synthetic control approach is arguably the most important innovation in the practical toolkit of policy evaluation in the last 15 years (Caliendo & Kopeinig, 2008; Smith & Todd, 2001).

3. DATA

The data are 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: To restrict our analysis 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).

None of these agencies have information on what other funders are doing. It has historically been a herculean task to manually link and standardize data across disaggregated systems. However, new data sources have become available, such as Dimensions⁶, a Digital Science database tool aggregating publications, citations, grants, clinical trial patents, and policy papers (Herzog, Hook, & Konkiel, 2020). Using advanced techniques such as 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 1 billion citations. The Dimensions team has also linked the INCa data to the Dimensions database and used the database to capture information on other cancer funders and researchers. This was achieved both by using data provided by INCa to the Dimensions team and by processing funding acknowledgment texts in research articles.

Funding agencies also typically do not have a way of tracing the research activity of scientists both before and after the award of a grant (i.e., the initial results of funding). To construct data on research output prior to funding, we worked closely with the ORCID⁷ 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 applications, publication submissions, peer review, and data set deposits. 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 U.S.

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.

⁷ <https://orcid.org/>

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 Research⁸.

The advantage of such automated approaches is that they can be quite cost effective for agencies. The use of ORCID has been extremely popular because it replaces manual and expensive approaches to populating institutional reporting systems to support national assessment programs. Similarly, Dimensions data can be used to replace the time-consuming manual effort of asking researchers to report on related grants. Finally, if an integrated database is established, it can substantially reduce the time to produce reports and analyses on the part of the funding agencies. On the researcher side, it has been estimated that the grant reporting burden takes up as much as 40% of a faculty member's time in the United States; that burden can be relieved with prepopulated information (Decker, Wimsatt, et al., 2007).

3.1. Data on Funded Research

In general, a major challenge with an effort such as this is that there is quite limited information on *what* research is funded. Individual agencies provide information on individual grants: For example, the European Union's CORDIS⁹, NSF's *research.gov*¹⁰, and the NIH reporter¹¹ are very useful tools for capturing information about individual awards, but do not provide a good overview of the funding landscape for research in particular fields or across agencies. It would be a huge task to pull data together from multiple sources and standardize the information.

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 system¹². 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 in the Dimensions platform. Using the coded grants as a training set, RCDC categories were assigned to all publications and grants in the Dimensions database. In addition, Dimensions provided more detailed cancer-specific codes, called Common Scientific Outline (CSO) codes, which were developed by the International Cancer Research Partnership¹³. Dimensions integrated INCa researchers into their database using automated approaches that were also manually validated.

3.2. Measuring Scientific Activity

The second major challenge is to trace the research activity of scientists both before and after the award of a grant. We were able to do this by tracing the activities of individuals through their ORCID identifier. The integration of 2007–2012 INCa data into the Dimensions database showed that only a minority of funded researchers possessed an ORCID identifier. INCa sent an email to all funded INCa researchers describing the HELIOS project and requested the researchers to click on a customized link to create an ORCID identifier, populate their ORCID record with publication and grant data, and share their ORCID identifier with the HELIOS team. This led to the sharing of 174 ORCID records; 749 researchers did not confirm their ORCID identifiers. Accounting for a

⁸ <https://orcid.org/organizations/funders/policies>

⁹ https://cordis.europa.eu/projects/home_en.html

¹⁰ <http://www.research.gov/>

¹¹ <http://projectreporter.nih.gov/reporter.cfm>

¹² <https://report.nih.gov/rcdc/>

¹³ <https://www.icrpartnership.org/>

number of invalid email addresses in the INCa 2007–2012 database, this represents a 20% response rate.

3.3. Construction of the Analytical Sample

Two data sets were provided by Dimensions, which serve to create a comparison group for our model: all prior and subsequent grants awarded to these researchers, and all publications assigned to the same researchers. The retained grant data included dollar amount, duration, and topic (RCDC names and codes). The publication data included title, journal, publication date, number of citations, and publication topic (CSO and RCDC names and codes).

The Dimensions database allowed us to collect all relevant grants attributed between 2007 and 2012 to the five funding agencies that we consider. We also collected all principal investigator names associated with these grants. We pulled all prior and subsequent grants and publications assigned to these researchers. The initial sample of INCa/INSERM/DGOS funded individuals is made up of 914 researchers, funded by 876 grants. For the four comparison agencies that were identified, we pulled all funded cancer grants and corresponding researchers. The distribution of the initial sample is presented in Table 1.

In the case of researchers who are funded by several comparison agencies over their career, only their first grant is considered. Finally, a researcher’s original affiliation is determined by querying the affiliation for every publication and keeping the earliest nonnull record. We are unable to impute the affiliation of researchers with no associated publications, or of researchers where all publications have no corresponding affiliation data. These researchers are dropped as well in the final analytical sample.

4. IMPACT MEASUREMENT

At a statistical level, the evaluation of a scientific hypothesis—such as 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 nonrandom participation and identify an appropriate comparison group (or groups). The standard impact evaluation framework is to determine the impact (Δ) or causal effect of a program (P) on an outcome of interest (Y):

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

In other words, the causal impact of a program (such as receiving a research grant from a prestigious funder) on a scientific outcome is the difference between the outcome with and

Table 1. Analytical sample

	Number of researchers	Number of researchers (%)	Number of grants	Number of grants (%)
Cancer Research UK	859	7.1	847	8.5
INCa/INSERM/DGOS	914	7.6	876	8.8
National Cancer Institute	8,521	70.5	7,244	73.0
NHMRC	1,543	12.8	785	7.9
Wellcome Trust	246	2.0	170	1.7

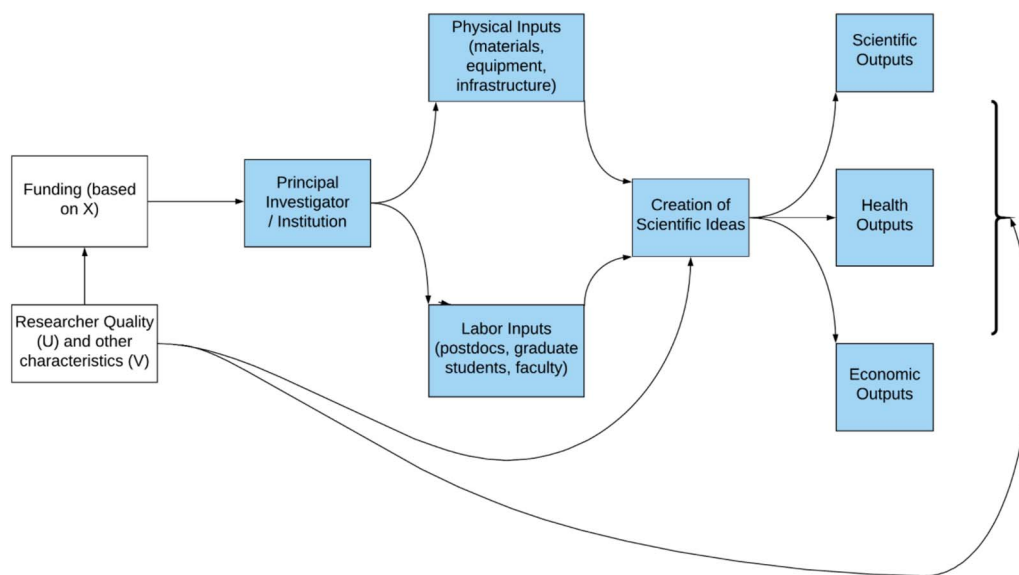


Figure 1. Conceptual framework.

without the program, in this case receiving a prestigious research award (Gertler, Martinez, et al., 2016). The framework is useful to describe outputs, confounding factors, and develop counterfactuals.

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, such as 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.

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, such as education, criminality, or employment. 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 ¹⁴. Positive selection is likely to upwardly bias estimates of the effect of science funding. For example, the estimated impact is likely to overstate the effect of the funding itself if higher ability researchers are more likely to be selected for funding by one of these agencies. That is, the best researchers would have been successful over time, regardless of whether they actually receive funding. Thus, naïve estimates of the effect of funding would conflate selection into receiving funding with the casual effect of receiving a research award.

¹⁴ This description is directly taken from Abadie and Cattaneo (2018).

4.1. Inverse Propensity Score Weighting

One popular approach to estimating impact is to construct comparison groups consisting of similar scientists as the focal group, but who either do not receive funding or receive funding from a different agency. This approach requires having reasonable measures of the confounding covariates, U and V from Figure 1. This approach has been extensively applied since its introduction by Rosenbaum and Rubin (1985), and essentially groups similar individuals based on the propensity to be treated (Caliendo & Kopeinig, 2008; Heckman, Ichimura, et al., 1998)¹⁵. There are several canonical approaches, which all rely on the researcher empirically estimating the likelihood that an individual is selected into treatment (i.e., the propensity of treatment). In the present context, we estimate the likelihood that an individual receives funding from a given agency based on observable characteristics obtained from the Dimensions data. To assess the impact of treatment (receiving funding from INCa), a comparison group consisting of untreated individuals (i.e., those who did not receive INCa funding) but who are empirically similar to the treated individuals. Some approaches would construct this control group using a one-to-one or one-to-many matching strategy based on selecting only control individuals with the highest propensity score. Here, we utilize all of the untreated individuals but weight the observations based on the inverse propensity score such that the control individuals who most resemble treated individuals are given more weight in the resulting estimation sample.

The efficacy of propensity scores is limited by the assumption that selection into treatment is based on observable characteristics. Clearly, cases exist when confounding covariates are not measurable based on available data. In this case, the canonical approach is to apply a “difference-in-difference” estimator where the confounding factors are assumed to be time invariant in levels within a given individual. The identifying assumption then depends only on equality of trends, rather than levels, in the pre period. This method permits the identification of change between two time periods (i.e., before and after the receipt of INCa funding). This methodology is particularly useful when it is not possible to directly observe a rich set of population (persons, firms, etc.) characteristics. Combining difference-in-differences with propensity scores is commonly referred to as a *synthetic control approach* and helps to mitigate the limitations of applying each approach independently.

4.2. Implementation, Comparisons with Inverse Propensity Score Weighting

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. Overall, our analysis accounts for 9,922 grants, awarded to 12,083 unique researchers between 2007 and 2012, of which 859 were from CRUK, 914 from INCa, 8,521 from NCI, 1,543 from NHMRC, and 246 from Wellcome Trust. As noted, we consider only a researcher’s first career grants, so each researcher is counted only once and associated with only one grant (subject to the reporting biases noted above). Thus, we exclude exceptional cases of researchers funded by several of the five agencies within a short period of their early career. As this study is observational and nonrandomized, there are differences in baseline characteristics between the researchers from the five comparison agency groups. As discussed, we address this issue using inverse propensity score weighting, which balances the groups and reduces baseline differences in characteristics across agencies.

Using the group of INCa/INSERM/DGOS researchers as our reference group, we weight the individuals to minimize the differences on a number of covariates between the focal and reference group. The covariates used to calculate the propensity scores are based on the first years of

¹⁵ A technical summary is provided in the supplementary information.

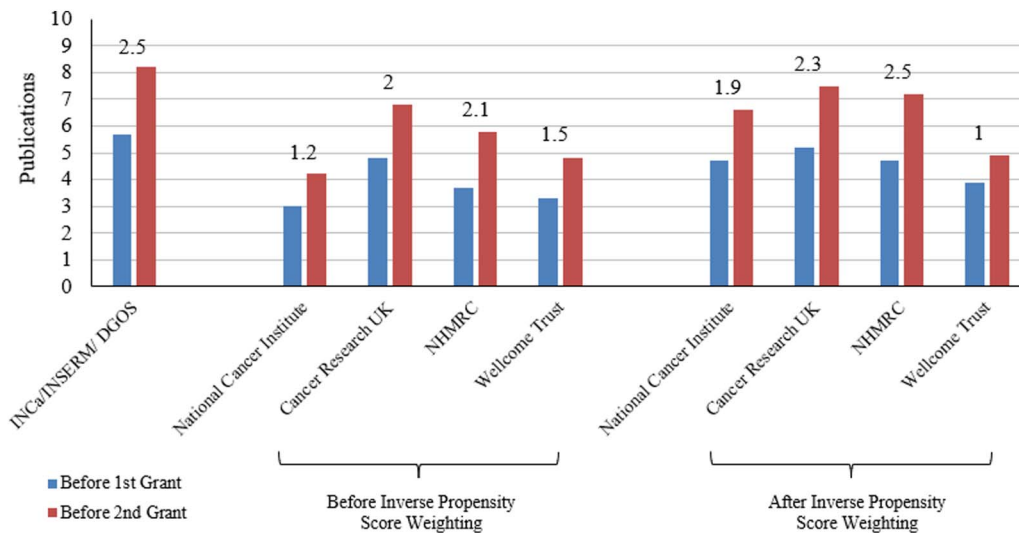


Figure 2. Number of publications.

available data on every researcher as well as time-invariant demographic characteristics. In particular, we include the number of publications by a given researcher in the first year of their career as well as their gender and a second-order polynomial of career age.¹⁶ We also include the RCDC and CSO codes of a researcher's first year of publications, which are used in the calculation of the propensity scores. All publications from a researcher's first career year are pulled using the Dimensions API, along with their RCDC and CSO topic codes. Based on these publications, researchers are assigned a distribution of topics for the first year of their research.

In this section, we examine several measures of scientific output before and after a maximum of 5 years after a researcher's first career grant is made. The measures of scientific output include total number of publications, mean citations per publication, number of unique collaborators, and the breadth of countries represented by collaborators. In each of the figures below, we present summary statistics from the raw unadjusted data by agency before and after the researcher receives their first career grant. We also include results where the comparison agencies have been adjusted using inverse propensity score weights such that the researchers who have characteristics closer to the characteristics of an INCa researcher have a greater sample weight. The supplementary information contains formal regression results corresponding with each table where we estimate panel data models that include time and research fixed effects as well as the variables included in the estimation of the propensity scores (i.e., so-called doubly robust estimation; see Bang & Robins, 2005).

Figure 2 presents the change in the number of publications before and after receiving funding from each agency for those whose first major research grant was from that agency. The left panel plots the mean number of publications 5 years before and after receiving the research award in the raw data. The right panel shows the same statistic for funders other than INCa, but where the observations have been weighted based on the inverse propensity score (i.e., based on how similar they are to INCa researchers). In both panels, the change in the number of publications is annotated above the bars. The change in the number of publications for INCa researchers (2.5) was larger relative to all of the comparison agencies. After reweighting the data based on the

¹⁶ Note that we also include a binary variable indicating whether an individual confirmed their ORCID in dimensions.

inverse propensity score, the right panel shows that only the NHMRC researchers had a similar change in the number of publications (2.5). Consistent with our expectation that selection into funding from INCa would bias naïve estimates upwards, our findings suggest that INCa has among the highest impact on research output but not as much after we adjust the estimates based on observable differences between the groups. Table S.1 in the supplementary information contains regression results from doubly robust estimation, where controls used to estimate the propensity score are again included in the outcome regression, in addition to affiliation and year fixed effects.

Figure 3 presents the change in the number of citations per publication before and after receiving funding from each agency for those whose first major research grant was from that agency. The left panel plots the mean impact per paper 5 years before and after receiving the research award in the raw data. The right panel shows the same statistic for funders other than INCa but where the observations have been weighted based on the inverse propensity score (i.e., based on how similar they are to INCa researchers). In both panels, the change in the number of publications is annotated above the bars. There is a decline in the impact per paper for INCa researchers (19) which was significantly smaller relative to the comparison agencies. After re-weighting the data based on the inverse propensity score, the gap between INCa and the comparison agencies shrinks, but INCa researchers still have the smallest decline in impact per paper. The variable was constructed by querying all publications by a given researcher in a given year and averaging the citation counts. As the only available citation count is the publications' current citation count, this variable reflects how often these publications have been cited since their issue, regardless of the publication year. Table S.2 in the supplementary information contains regression results from doubly robust estimation, where controls used to estimate the propensity score are again included in the outcome regression in addition to affiliation and year fixed effects.

Figure 4 presents the change in the number of unique collaborators before and after receiving funding from each agency for those whose first major research grant was from that agency. The left panel plots the number of unique collaborators 5 years before and after receiving the research award in the raw data. The right panel shows the same statistic for funders other than

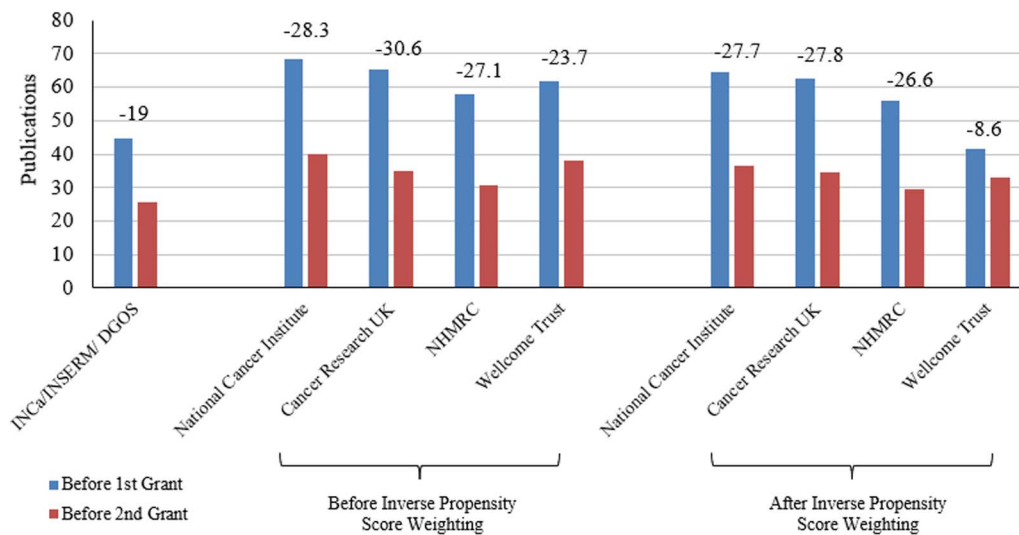


Figure 3. Mean citations per publication.

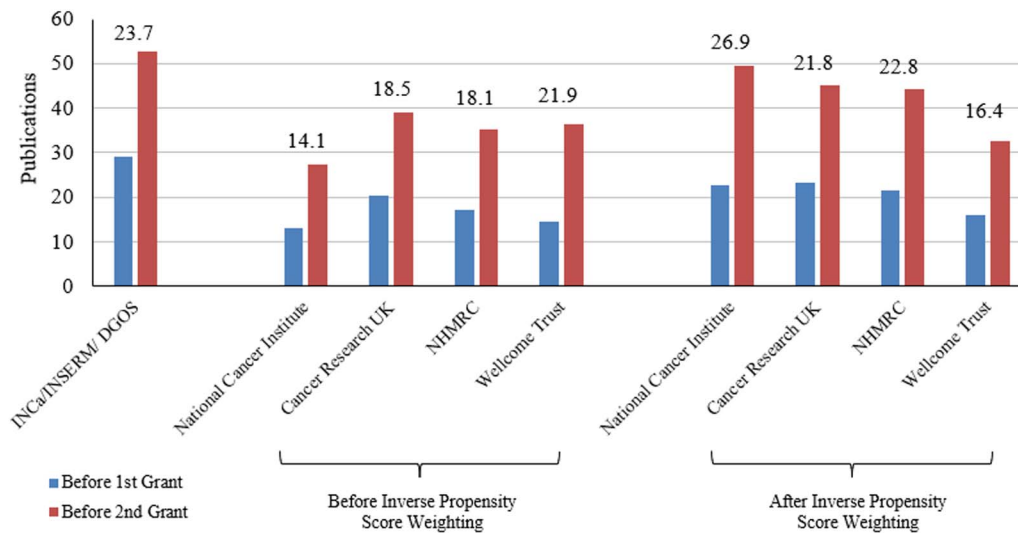


Figure 4. Number of collaborators.

INCa, but where the observations have been weighted based on the inverse propensity score (i.e., based on how similar they are to INCa researchers). In both panels, the change in the number of publications is annotated above the bars. The change in the number of collaborators for INCa researchers (23.7) was larger relative to all of the comparison agencies. After reweighting the data based on the inverse propensity score, the right panel shows that only the National Cancer Institute researchers had a similar change in the number of publications (26.9). Consistent with our expectation that selection into funding from INCa would bias naïve estimates upwards, our findings suggest that INCa has among the highest impact on the number of collaborators, but not as much after adjusting the estimates based on observable differences between the groups. This variable was constructed by querying all publications by a given researcher in a given year, and all authors associated with these publications. Distinct researchers are counted using their unique internal Dimensions ID. Table S.3 in the supplementary information contains regression results from doubly robust estimation, where controls used to estimate the propensity score are again included in the outcome regression in addition to affiliation and year fixed effects.

Figure 5 presents the change in the number of distinct countries where the author has collaborators before and after receiving funding from each agency for those whose first major research grant was from that agency. The left panel plots the number of unique countries 5 years before and after receiving the research award in the raw data. The right panel shows the same statistic for funders other than INCa, but where the observations have been weighted based on the inverse propensity score (i.e., based on how similar they are to INCa researchers). In both panels, the change in the number of publications is annotated above the bars. The change in the number of distinct countries for INCa researchers (1.4) was larger relative to all of the comparison agencies. After reweighting the data based on the inverse propensity score, the right panel shows that only the National Cancer Institute researchers had a similar change in the number of publications (1.6). Again, consistent with our expectation that selection into funding from INCa would bias naïve estimates upwards, our findings suggest that INCa has among the highest impact on the number of collaborators, but not as much after we adjusting the estimates based on observable differences between the groups. This variable is constructed in a fashion similar to the number of collaborators, by querying all of the publications by a researcher in a given year, and the countries of affiliation of all the coauthors

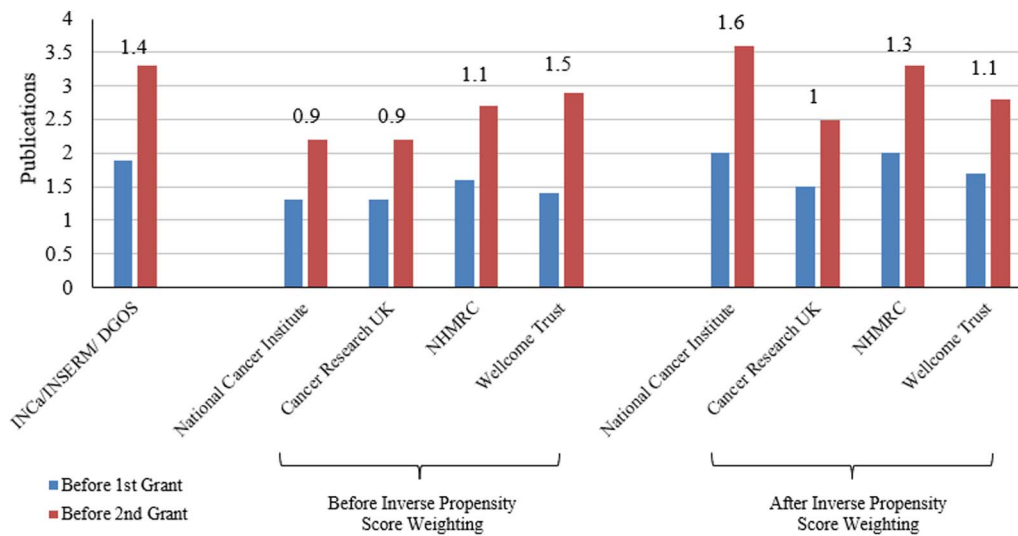


Figure 5. Number of collaborations in different countries.

associated with these publications. It reflects the number of distinct countries among a researcher’s coauthors. Table S.4 in the supplementary information contains regression results from doubly robust estimation, where controls used to estimate the propensity score are again included in the outcome regression in addition to affiliation and year fixed effects.

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, while INCa specialized 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* (Benavente, Crespi, et al., 2012; Bronzini & Iachini, 2014). Not all biases are accounted for, of course. There may be systemic differences in the quality of reporting across funders. For example, NCI does not require an ORCID, which may result in lower quality information about the NCI funded researchers’ publications. The Matthew effect may also be a factor (Bol, de Vaan, & van de Rijt, 2018).

5. RECOMMENDATIONS FOR FUNDERS

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 by linking the funding that goes to individuals with their subsequent scientific activity. 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 and compare their activities across funding sources—a science of science. It applies statistical techniques to the cutting edge of the program evaluation literature to examine the relative impact of French funding for cancer research.

The results presented in this study were designed to be illustrative of a use case in France but also to lay the foundations for a scalable approach. We show that it is not only possible but also low cost to design a data system that can make comparisons across scientific agencies and apply statistical techniques that can control for both observed and unobserved factors. The design in this paper traces the activities of individuals subsequent to receiving research funding and their scientific activities in terms of publications, citations, collaborations, and international activity. Much more can be done with this approach. With thorough statistical methods at hand, and collected data, funders could agree to share consistent information about who is funded, the funding amounts, and descriptions of the funding investments. In tandem, the scientific community could develop more outcome measures, such as student placements, data and code sharing activities, and 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 (Lane et al., 2015). That work has been coupled with the Innovation Measurement Initiative at the U.S. Census Bureau and has led to deeper understanding of how research activity stimulates economic and scientific innovation (Lane et al., 2015; Teich, 2018). In Australia, a new recommendation from the Australian Parliament calls for the use of ORCID and streamlined reporting to enable evaluation (Australian Parliament, 2018). 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 (Bosman, 2018). 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.

There are also limitations, as with any research. There was limited information about the research team on each project, so it was impossible to identify the contributions of individuals other than the principal investigator. The European funding system does not publish the names of the principal investigators and there was also no information on the contributions and capacity of the institutions to which the researchers belonged. Further research would include such factors to get even deeper insights into the contribution of research funding to the productivity of an individual researcher.

ACKNOWLEDGMENTS

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.

AUTHOR CONTRIBUTIONS

G rard Chevalier: Conceptualization, Data analysis, Writing—original draft, Writing—review & editing. Christine Chomienne: Conceptualization. Nicolas Guetta Jeanrenaud: Data analysis, Writing—original draft, Writing—review & editing. Matthew Ross: Data analysis, Writing—original draft, Writing—review & editing. Julia Lane: Data analysis, Writing—original draft, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

This work was supported through funds provided by Plan Cancer 2014–2019 (INCa-ITMO Cancer).

DATA AVAILABILITY

The data sets generated during and/or analyzed during the current study are not publicly available due to confidentiality clauses, but are available from the corresponding author on reasonable request.

All codes used to generate results presented in this study are available on GitHub: <https://github.com/nico-gj/helios>.

REFERENCES

- Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Econometrics*, 10, 465–503. **DOI:** <https://doi.org/10.1146/annurev-economics-080217-053402>
- Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *J. Econ. Perspect.* 31, 3–32. **DOI:** <https://doi.org/10.1257/jep.31.2.3>
- Australian Parliament. (2018). *Australian Government Funding Arrangements for non-NHMRC Research*. https://www.aph.gov.au/Parliamentary_Business/Committees/House/Employment_Education_and_Training/FundingResearch/Report/section?id=committees/reportrep/024212/26656
- Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4), 962–973. **DOI:** <https://doi.org/10.1111/j.1541-0420.2005.00377.x>, **PMID:** 16401269
- Benavente, J., Crespi, G., Garone, L., & Maffioli, A. (2012). The impact of national research funds: A regression discontinuity approach to the Chilean FONDECYT. *Research Policy*, 41(8), 1461–1475. **DOI:** <https://doi.org/10.1016/j.respol.2012.04.007>
- Bernanke, B. S. (2011). Promoting research and development: The government's role. *Issues in Science and Technology*, 27, 37–41.
- Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, 115(19), 4887–4890. **DOI:** <https://doi.org/10.1073/pnas.1719557115>, **PMID:** 29686094, **PMCID:** PMC5948972
- Bosman, J. (2018). Innovations in scholarly communication: Changing research workflows. Available at https://www.academia.edu/12317222/101_Innovations_in_Scholarly_Communication_the_Changing_Research_Workflow
- Bozeman, B., Dietz, J. S., & Gaughan, M. (2001). Scientific and technical human capital: An alternative model for research evaluation. *International Journal of Technology Management*, 22(7/8), 716–740. **DOI:** <https://doi.org/10.1504/IJTM.2001.002988>
- Bozeman, B., & Rogers, J. D. (2002). A churn model of scientific knowledge value: Internet researchers as a knowledge value collective. *Research Policy*, 31(5), 769–794. **DOI:** [https://doi.org/10.1016/S0048-7333\(01\)00146-9](https://doi.org/10.1016/S0048-7333(01)00146-9)
- Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, 9, 439–468. **DOI:** <https://doi.org/10.1093/jeg/lbp008>
- Bronzini, R., & Iachini, E. (2014). Are incentives for R&D effective? Evidence from a regression discontinuity approach. *American Economic Journal: Economic Policy*, 6, 100–134. **DOI:** <https://doi.org/10.1257/pol.6.4.100>
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31–72. **DOI:** <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Corrado, C., Haskel, J., & Jona-Lasinio, C. (2017). Knowledge spillovers, ICT and productivity growth. *Oxford Bulletin of Economics and Statistics*, 79(4), 592–618. **DOI:** <https://doi.org/10.1111/obes.12171>
- Corrado, C., & Lane, J. (2009). Using cyber-enabled transaction data to study productivity and innovation in organizations. *Global COE Hi-Stat Discussion Paper Series 099*. The Conference Board.
- Decker, R., Wimsatt, L., Trice, A., & Konstan, J. (2007). *A profile of federal-grant administrative burden among Federal Demonstration Partnership faculty*. A Report of the Faculty Standing Committee of the Federal Demonstration Partnership.
- Gertler, P., Martinez, S., Premand, P., Rawlings, L., & Vermeersch, C. (2016). *Impact evaluation in practice*. The World Bank. **DOI:** <https://doi.org/10.1596/978-1-4648-0779-4>
- Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5), 1017–1098. **DOI:** <https://doi.org/10.2307/2999630>
- Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down the barriers between scientometricians and data. *Quantitative Science Studies*, 1(1), 387–395. **DOI:** https://doi.org/10.1162/qss_a_00020
- Jaffe, A., & Jones, B. (2014). *The changing frontier: Rethinking science and innovation policy*. Chicago, IL: University of Chicago Press. **DOI:** <https://doi.org/10.7208/chicago/9780226286860.001.0001>
- Joly, P.-B., et al. (2015). ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. *Research Evaluation*, 24(4), 440–453. **DOI:** <https://doi.org/10.1093/reseval/rvv015>
- Lane, J., Owen-Smith, J., Rosen, R., & Weinberg, B. (2015). New linked data research investments: Scientific workforce, productivity, and public value. *Research Policy*, 44(9), 1659–1671. **DOI:** <https://doi.org/10.1016/j.respol.2014.12.013>, **PMID:** 26335785, **PMCID:** PMC4553239
- MacIlwain, C. (2010). Science economics: What science is really worth. *Nature*, 465, 682–684. **DOI:** <https://doi.org/10.1038/465682a>, **PMID:** 20535177
- Mairesse, J., & Mohnen, P. (2010). In B. Hall & N. Rosenberg (Eds.) *Handbook of the Economics of Innovation, Vol. II*. New York, NY: Academic Press.
- Mairesse, J., Mohnen, P., & Kremp, E. (2005). The importance of R&D and innovation for productivity: A reexamination in light of the

- French innovation survey. *Annals of Economics and Statistics*, Issue 79/80. **DOI:** <https://doi.org/10.2307/20777586>
- Marburger, J. (2005). Wanted: Better benchmarks. *Science*, 308(5725), 1087. **DOI:** <https://doi.org/10.1126/science.1114801>, **PMID:** 15905367
- Martin, B. R. (2011). The Research Excellence Framework and the “impact agenda”: Are we creating a Frankenstein monster? *Research Evaluation*, 20, 247–254. **DOI:** <https://doi.org/10.3152/095820211X13118583635693>
- National Science and Technology Council. (2008). *The science of science policy: A federal research roadmap*. Washington, DC: Science of Science Policy Interagency Task Group.
- Penfield, T., Baker, M. J., Scoble, R., & Wykes, M. C. (2014). Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation*, 23(1), 21–32. **DOI:** <https://doi.org/10.1093/reseval/rvt021>
- Powell, W., & Giannella, E. (2010). In *The handbook of innovation*. Amsterdam: Elsevier. **DOI:** [https://doi.org/10.1016/S0169-7218\(10\)01013-0](https://doi.org/10.1016/S0169-7218(10)01013-0)
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5), S71–S102. **DOI:** <https://doi.org/10.1086/261725>
- Rosenbaum, P. R. & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33–38. **DOI:** <https://doi.org/10.1080/00031305.1985.10479383>, <https://doi.org/10.2307/2683903>
- Smith, J. A., & Todd, P. E. (2001). Reconciling conflicting evidence on the performance of propensity-score matching methods. *American Economic Review*, 91(2), 112–118. **DOI:** <https://doi.org/10.1257/aer.91.2.112>
- Spaapen, J., & Van Drooge, L. (2011.) Introducing “productive interactions” in social impact assessment. *Research Evaluation*. 20(3), 211–218. **DOI:** <https://doi.org/10.3152/095820211X12941371876742>
- Teich, A. H. (2018). In search of evidence-based science policy: From the endless frontier to SciSIP. *Annals of Science and Technology Policy*, (2)2, 75–199. **DOI:** <https://doi.org/10.1561/110.00000007>
- Whittington, K. B., Owen-Smith, J., & Powell, W. W. (2009). Networks, propinquity, and innovation in knowledge-intensive industries. *Administrative Science Quarterly*, 54, 90–122. **DOI:** <https://doi.org/10.2189/asqu.2009.54.1.90>