Scientometrics (2013) 94:1253–1273
DOI 10.1007/s11192-012-0841-3

Effects of large-scale research funding programs: a Japanese case study

Takanori Ida · Naomi Fukuzawa

Received: 7 April 2012 / Published online: 9 September 2012

Abstract This study investigates the effects of large-scale research funding from the Japanese government on the research outcomes of university researchers. To evaluate the effects, we use the difference-in-differences estimator and measure research outcomes in terms of number of papers and citation counts per paper. Our analysis shows that the funding program led to an increase in the number of papers in some fields and an increase in the citation counts in the other fields. A comparison of our estimation results with assessment data obtained from peer reviews showed important differences. Since the characteristics of research vary according to the field, bibliometrics analysis should be used along with the peer review method for a more accurate analysis of research impact.

Keywords Research assessment · Difference-in-differences · Government grants · University research · Bibliometrics · Peer review

Introduction

In the report titled "Present Status of Research Evaluations and its Future in Japan", the Science Council of Japan has indicated that research evaluations play an important role in ensuring accountability in the use of research funds as well as in promoting research activities and improving their quality (Committee for Research Evaluations 2008). However, the same report has also highlighted that the research evaluation system of the Japanese government is not adequately developed and that research evaluation in Japan is a complex process, requiring considerable thought and effort from the evaluators and the evaluated academic units. In the United Kingdom, allocation of research grants to universities is based on the results of research evaluation process, known as the Research Assessment Exercise (RAE), which has already been undertaken six times. Although the RAE offers a prototype for evaluation methodology, it requires substantial inputs in terms

T. Ida · N. Fukuzawa (✉)
Graduate School of Economics, Kyoto University, Yoshida, Sakyo, Kyoto 606-8501, Japan
e-mail: fukuzawa.naomi.62w@st.kyoto-u.ac.jp

of time and money (Kostoff 1994; Oppenheim 1995). The Research Excellence Framework (REF), which will be completed in 2014 and will replace the RAE, will assess the quality of research in the three assessment areas: research output (65 %), impact (20 %), and environment (15 %); it will be carried out using bibliometric indicators such as papers and citation data from Scopus (Elsevier).

The purpose of this paper is to critically assess the performance of a large-sized research funding program in Japan, known as the 21st Century Centers of Excellence (COE) Program, using a different evaluation technique, a bibliometric analysis. The aim of this funding program was to cultivate a competitive academic environment among Japanese universities by providing targeted support for the creation of world-class research and education bases. Nearly 6,000 researchers benefited from the COE program; however, their publication-related achievements have not been captured in a database. Institutions that applied for funds through the COE included PhD departments of graduate schools and university-affiliated research institutes. Applications for the COE program were reviewed by a team of 1,000 referees, each a specialist or leading authority in the subject field. Category-specific subcommittees evaluated the proposals through hearings and panel interviews. Applications were shortlisted by a screening committee, and the selection results were submitted to the program committee for final judgment. Applications for the COE program, which awarded research funding for a 5-year term, were accepted in FY 2002 (5 fields), FY 2003 (5 fields), and FY 2004 (1 field).[1] Program performance was assessed by subject-specific subcommittees using peer review methods at a 2 year interval to verify the progress of the funded research projects, and final assessments were carried out after project completion.

In this study, we attempt to evaluate the effects of the COE program through bibliometric analysis, which complements peer review methods. To our knowledge, this is the first study that intends to evaluate the success of the COE program by analyzing its effects on the publication-related achievements of its recipients. Our analysis covers research achievements in eight fields recognized under the COE program: (1) life sciences, (2) information sciences, electrical and electronic engineering (henceforth information sciences), (3) chemistry and material sciences, (4) humanities, (5) social sciences, (6) medical sciences, (7) mechanical, civil, architectural and other fields of engineering (henceforth mechanical engineering) and (8) mathematics, physics, and earth sciences (henceforth the mathematics and physics). Three interdisciplinary fields were excluded from the study as the research achievements in these fields were difficult to evaluate.

Data for bibliometric analysis of research outcomes is typically obtained from scientific databases. To construct the outcome indexes, we used the number of papers published by each researcher and the number of citations attributed to the researcher. We then employed the difference-in-differences (DID) estimator to evaluate the difference between "the before–after outcomes in the treatment group, which received funding" and "the before–after outcomes in the control group, which did not receive funding." Thus, we controlled for a certain number of selection biases that stemmed from nonrandom assignment of the COE program.

Although, other studies have attempted to analyze the effects of funding programs with the help of control and treatment groups or with a before–after comparison, none of them have used DID estimation, which combines both treatment–control comparison and a before–after comparison to measure the effects of the treatment. Gaughan and Bozeman

[1] Details are given on the official site of the 21st COE Program (http://www.jsps.go.jp/english/e-21coe/index.html) of the Japan Society for the Promotion of Science.

(2002) examined the effects of funding from the National Science Foundation (NSF) on the number of papers published. They found no significant difference in the total number of papers published between the treatment and control groups. Gaughan and Ponomariov (2008) compared their treatment group, consisting of researchers who received center support from the National Institute of Child Health and Human Development (NICHD), with a suitable control group and showed that the center affiliation did not have a significant influence on the publication rate. Gaughan (2009) compared the treatment and control groups of a training program at the National Institute of Health (NIH) and showed that center affiliation did not have a significant impact on the paper publication rate adjusted by age. In a separate study, Bozeman and Gaughan (2007) showed that researchers who received industry grants or government grants had a higher involvement rate in the industry. The number of industry and government grants also had a significantly positive effect on industry involvement. Dietz and Bozeman (2005) compared the 5 year mean publication rate of researchers who shifted from universities to industries and vice versa, and showed that the shift had an important effect on productivity.

Caution must be exercised when interpreting the increased effects in different academic disciplines. In the natural sciences, the time lag between publication and citation is considered to be short, therefore the effects of research funding are comparatively easy to ascertain. As a result of calculating a proceedings ratio in the information sciences fields (such as information sciences, telecommunications engineering, computer sciences, media science, bioinformatics engineering, and semiconductor electronics which are included in the researcher's department of this fields in the COE program), there exists a stronger tendency to publish proceedings rather than journal articles. In social sciences and the humanities, the research cycle is longer, and the number of citations is lower. Thus, the field or discipline is a key factor affecting research evaluations.

This paper is organized as follows. "Methods" describes the indexes of research outcomes and the DID estimator method. "Data" provides details on the scientific database used for the study and the method employed for selecting data. "Estimation models" explains the truncation biases of the citation counts and the DID estimator model. "Life sciences" describes the estimation results. The lase section presents "Discussion and conclusion".

Methods

In this section, we explain the indexes of research outcomes and the methods used for evaluating the effectiveness of a research support program.

Indexes of research outcomes

Peer review, as a method of research evaluation, is a costly and lengthy process and suffers from subjective biases, stemming from the "old boy" networks that exist in established fields and "halo" effects, which refer to the higher likelihood of funding for more prominent scientists (Gibbons and Georghiou 1987; Kostoff 1994; Oppenheim 1995).

Many studies have investigated the correlation between peer review and bibliometrics (Anderson et al. 1978; Zhu et al. 1991; Oppenheim 1995, 1997; Rinia et al. 1998, 2001). For example, Oppenheim (1995) found a statistically significant positive correlation between the number of citations received by a department in total and the RAE rating of that department, derived from the peer-review method. In another study Van Raan (2006) using the results of an evaluation study of 147 university chemistry research groups in the

Netherlands during the period between 1991 and 2000, showed that the h-index and bibliometric indicators are positively correlated. Opthof and Leydesdorff (2011) used Van Raan's data and showed that a 5-point peer rating scale was uncorrelated with citations per paper/mean field citation score (CPP/FCSm) and h-index. Furthermore, they showed that "none of the citation-based indicators was able to discriminate between the categories 'good' and 'excellent' that were distinguished by the peer review".

Although each discipline of study is unique, the number of citations of patents is often used as an objective evaluation index for measuring patent value. Jaffe et al. (2000), Harhoff et al. (1999), and Lanjouw and Schankerman (1999) showed a positive correlation between the number of citations received by a patent and its importance (both economic and technological), as well as its quality and value. In line with the above studies, we use the number of papers as an activity index, while the citation counts reflect the quality of scientific research. Although the types of academic publication contained in the database included articles conference papers, reviews, and letters, we restricted our analysis to only research articles, as these have passed through the most rigorous peer review.

Treatment effect analysis

The average treatment effect (ATE) is one of the methods of evaluating the effects of a research funding program (Rosenbaum and Rubin 1983). ATE is the average difference between the outcome with treatment y_1 and the outcome without treatment y_0. It is defined as follows:

$$\mathrm{ATE} = E(y_1 - y_0) \tag{1}$$

Let the variable T be a binary treatment indicator, where $T = 1$ denotes the treatment group and $T = 0$ denotes the control group. An important assumption in order to identify treatment effects is the conditional independence assumption or the strongly ignorable treatment assignment (Rosenbaum and Rubin 1983), which states that conditional on a vector of observable variables x, the outcomes are independent of the treatment. The conditional independence assumption is written as follows:

$$(y_1, y_0) \coprod T|x \tag{2}$$
$$0 < p(T = 1|x) < 1$$

where $\coprod$ represents independence, following Dawid (1979), and the ATE is defined as follows:

$$\begin{aligned} \mathrm{ATE} &= E(y_1 - y_0|x) \\ &= E(y_1|x) - E(y_0|x) \\ &= E(y_1|x, T = 1) - E(y_0|x, T = 0) \end{aligned} \tag{3}$$

Moreover, when the condition expressed in Eq. 2 holds, the average treatment effect on the treated (ATET) (Heckman and Robb 1985) is defined by Eq. 4.

$$\begin{aligned} \mathrm{ATET} &= E(y_1 - y_0|x, T = 1) \\ &= E(y_1|x, T = 1) - E(y_0|x, T = 1) \\ &= E(y_1|x, T = 1) - E(y_0|x, T = 0) = \mathrm{ATE} \end{aligned} \tag{4}$$

This is the expected value of the difference between the results of the treatment and the results without the treatment in the treatment group.

We thus estimate ATET, when the treatment of the program involves random assignment. However, selection biases are inevitable in this case, because the COE program is not based on random assignment. For this reason, we use the DID estimator, explained in the following subsection.

The difference-in-differences estimator

The DID estimator is the difference between "the average change in y in the treatment group over the course of the program" and "the average change in y in the control group over the same time (Stock and Watson 2007)". We can disregard the unobservable individual-specific characteristics by comparing the same individual's research achievements before and after. By removing the time effects, we can measure only the effects of the treatment (Cameron and Trivedi 2005; Lee 2005; Stock and Watson 2007; Wooldridge 2002). Let "Before" and "After" represent the states before and after the program, respectively, and let "Treatment" and "Control", respectively refer to the treatment group and the control group. The DID estimator can be written as in Eq. 5 and represented as in Fig. 1.

$$\begin{aligned} \mathrm{BA}^{\mathrm{Treatment}} &= \{E(y|T=1,\mathrm{After}) - E(y|T=1,\mathrm{Before})\} \\ \mathrm{BA}^{\mathrm{Control}} &= \{E(y|T=0,\mathrm{After}) - E(y|T=0,\mathrm{Before})\} \\ \mathrm{DID} &= \mathrm{BA}^{\mathrm{Treatment}} - \mathrm{BA}^{\mathrm{Control}} \end{aligned} \tag{5}$$

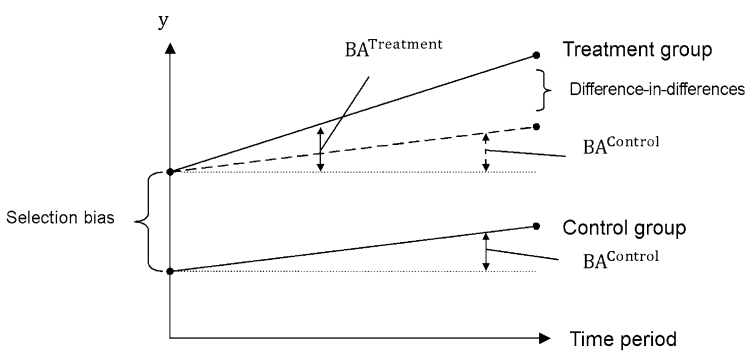
Let y_{b} be the outcomes before the program, and let $y_{1\mathrm{a}}$ be the outcomes after the program for the treatment group. Let $y_{0\mathrm{a}}$ be the outcomes after the program for the non-treatment group. Then, the DID estimator can be written as follows:

$$\mathrm{DID} = E(y_{1\mathrm{a}} - y_{\mathrm{b}}|T=1) - E(y_{0\mathrm{a}} - y_{\mathrm{b}}|T=0) \tag{6}$$

Using Eq. 4, we can write

$$\mathrm{ATET} = E(y_{1\mathrm{a}} - y_{0\mathrm{a}}|T=1)$$

When estimating ATET as the DID estimator, we assume the same time-effect condition (Lee 2005) the effects of not belonging to the treatment group are the same between the treatment group and the control group.



Source: We created this figure based on Stock and Watson(2007), p.482.

Fig. 1 The difference-in-differences estimator

$$E(y_{0a} - y_b | T = 1) = E(y_{0a} - y_b | T = 0) \quad (7)$$

Then we have

$$\mathrm{DID} = E(y_{1a} - y_b | T = 1) - E(y_{0a} - y_b | T = 0) = E(y_{1a} - y_{0a} | T = 1) = \mathrm{ATET}.$$

We measured the program effects assuming Eq. 7 holds.[2] In this study, the control group was formulated by extracting researchers randomly from the same university, the same graduate school, and the same major as the ones in the treatment group. Further, the indirect effects of belonging to the selected university were offset by using the DID estimator, which can estimate only the direct effects.

Data

In this section, we describe the scientific database used in this study and explain the process of obtaining data.

Database selection

A key concern in this analysis was identifying researchers correctly using the search function in the databases. Scopus assigned author IDs to prominent researchers, facilitating full name searches.[3] Web of Science (WoS) and Google Scholar are two other databases; previous research compares databases within a specific field. For example Meho and Yang (2007) compared the ranking of the citation counts (published between 1996 and 2005) between WoS and Scopus for 15 faculty members of the School of Library and Information Science at Indiana University, Bloomington, and found that the overall relative ranking of the faculty members did not change significantly between the two data sources. Norris and Oppenheim (2007) compared four scientific databases (WoS, Scopus, Google Scholar, and CSA Illumina) by using data from the RAE (2001) and from the International Bibliography in the Social Sciences, which includes non-English journals in the field. They concluded that Scopus offered the best coverage and could replace WoS as a tool to evaluate research impact in the social sciences fields.

Data on research outcomes

The author ID function on the Scopus database was not completely error-free leading to problems, especially in identifying authors whose affiliations had changed. To solve this problem, we first compiled a list that included details of the researchers' affiliations and departments at the start of the COE program, from the official COE website. We then searched for published papers and citations counts of each researcher.[4] To avoid problems in data accuracy, the collected data were filtered by matching their affiliations and departments listed in the published papers to those at the start of the COE program. Since

[2] This condition is weaker than the random assignment condition in Eq. 2.

[3] A drawback of the Scopus database is that it includes citation counts only since 1996. However, since we have used data only after 1997, this drawback did not affect our analysis.

[4] Although, we do not use the papers and citations that are not included in the source documents in Scopus, the ratio of papers not contained in Scopus is equal between the treatment and control groups within a field; these are offset by the DID estimation, which does not influence the estimation results.

this process could lead to the accidental omission of researchers who had changed their affiliations a number of times, we contacted each researcher via e-mail address and confirmed the list of publications with each researcher (researcher's self-check). Added to this, we excluded researchers who had enrolled, but left the COE program within a 5-year term.

To form the control group (researchers not engaged in the COE program), we randomly chose researchers from each university selected in the program. The ratios of full professors, divisions, and majors between the treatment and control groups were equalized. The number of teams and researchers used for analysis and the results of the researcher's self-check are summarized in Table 1.

To avoid any selection biases resulting from the researcher's self-check, we analyzed the difference between the "before–after outcomes of researchers who replied to self-check e-mail" and "before–after of the outcomes of researchers who did not reply to the self-check e-mail" using a t test of the number of papers. Because the null hypothesis of no difference could not be rejected in all fields at a 5 % level of statistical significance, we conclude that the biases are not serious.

Estimation models

Truncation bias of the citation counts

One of the major problems associated with the use of citation counts is truncation bias. The decline in the citations of recent papers is one of the causes of such truncation bias. To address this problem, some studies have used normalization, which is also used with forward citation of patents (Jaffe and Trajtenberg 1996; Jaffe and Lerner 2001; Hall et al. 2000, 2001). There are two approaches of normalization: the fixed-effects approach and the quasi-structural approach (Hall et al. 2001). The fixed-effects approach assumes normalization by dividing each citation by the corresponding year-field citation mean. The quasi-structural approach attempts to distinguish the multiple effects on citation via econometric estimation.

In this paper, we estimated the fixed effects using an eclectic mix of the two above-mentioned methods. Let the dependent variable be the citation counts per paper. We reviewed citation data from 1997 to 2007 for the fields funded by the COE program in FY 2002, and data from 1998 to 2008 for the fields funded by the COE program in FY 2003. We use the year dummy variable as an independent variable.[5] The regression equation is shown below:

$$\mathrm{Cited}_{it}^{\mathrm{p}} = \alpha + \beta d_{it}^{\mathrm{year}} + u_{it} \tag{8}$$

$\mathrm{Cited}_{it}^{\mathrm{p}}$: citation counts per paper for year t of researcher i

d_{it}^{year}: year dummy variable with some base year.

We used the least-squares method for each team of the treatment group and the control group. We subtracted the "estimated amount in each year" from the "citation counts in each year", because this estimated amount is negative, given that there is underestimation by truncation.

[5] We used 1998 as the base year for the fields incorporated in the COE program of FY 2002, and 1999 for the fields incorporated in the COE program of FY 2003.

Table 1 The number of researchers, teams, and the result of the researcher's self-check

	Adopted in FY 2002				Adopted in FY 2003			
	Life sciences	Information sciences	Chemistry, material sciences	Humanities	Social sciences	Medical sciences	Mechanical engineering	Mathematics and physics
The number of teams	28	20	21	20	25	35	22	23
The number of researchers in the treatment group	419	360	389	316	441	510	333	397
The number of researchers in the control group	109	102	102	93	125	137	98	109
Sent an e mail (proportion)	282 (67 %)	263 (73 %)	254 (66 %)	14 (15 %)	107 (49 %)	264 (53 %)	163 (58 %)	227 (70 %)
Obtained a reply (proportion)	163 (58 %)	127 (48 %)	110 (43 %)	9 (64 %)	62 (58 %)	145 (55 %)	68 (42 %)	113 (50 %)

Difference-in-differences estimator models

We conducted a linear panel regression analysis using 10 years of data for each researcher, excluding the year of program adoption. The regression model is shown below.

$$\mathrm{Paper}_{it} = \gamma_0 + \gamma_1 d_i^{\mathrm{treat}} + \gamma_2 d_{it}^{\mathrm{After}} + \delta d_i^{\mathrm{treat}} d_{it}^{\mathrm{After}} + \varepsilon_{it} \tag{9}$$

$$\mathrm{Cited}_{it}^{\mathrm{adjusted}} = \gamma_0' + \gamma_1' d_i^{\mathrm{treat}} + \gamma_2' d_{it}^{\mathrm{After}} + \delta' d_i^{\mathrm{treat}} d_{it}^{\mathrm{After}} + \varepsilon_{it} \tag{10}$$

Paper_{it}: total number of papers in year t of researcher i
$\mathrm{Cited}_{it}^{\mathrm{adjusted}}$: adjusted number of citations per paper per year of researcher i
d_i^{treat}: dummy variable indicating the researcher selected by the program
d_{it}^{After}: dummy variable of the time period after the program.

The DID estimator of the number of papers and citation counts is δ in Eq. 9 and δ' in Eq. 10, respectively.

Estimation results

In this section, we describe the DID estimation results. The DID estimation results for the eight science fields analyzed in this study are shown in Tables 2 and 3. Results of the average increased effects in the number of papers and citations are shown in Tables 4 and 6. After performing a Hausman's specification test, we were not able to reject the null hypothesis that "there is no correlation between the individual characteristics effect and the independent variables" in all the fields. Consequently, we adopted a random-effects model. We explain the results in detail below.

Life sciences

In the fields of life sciences, we found a statistically significant increase in the number of papers and citations. The main results are given as below.

- The number of papers per researcher per year increased from 3.08 to 4.38 after the introduction of the COE program. The COE effect accounted for 0.71 of total increase (Δ1.30), which was statistically significant at the 1 % level. Moreover, the citation counts per paper increased from 28.56 to 37.39 after the COE program (Δ8.83). The COE effect amounted to 4.67, which was also statistically significant at the 1 % level.
- We obtained statistically significant results for the number of papers authored by 13 teams out of a total of 28 teams. We also obtained statistically significant results for the citation counts of five teams.
- The effects of the COE program on national and public universities (22 teams) were compared with those on private universities (6 teams). The number of papers per researcher per year increased by 0.7 and 0.1, respectively. Similarly, citation counts per paper increased by 6.4 and 1.5, respectively (see Table 4).
- With regard to specific disciplines, biological sciences witnessed the highest increase in the number of papers, while the disciplines of life science and biological mechanisms and functions witnessed the highest increase in citation counts (see Tables 5, 6).
- We compared the results of the peer review methods with our estimation results. The three teams that had obtained the highest peer review assessment were among the 13

Table 2 The DID estimation results for the number of papers per researcher per year

	Adopted in FY 2002				Adopted in FY 2003			
	Life sciences	Information sciences	Chemistry, material sciences	Humanities	Social sciences	Medical sciences	Mechanical engineering	Mathematics and physics
Treatment group								
Before the COE	3.08	1.75	4.57	0.12	0.15	4.22	0.87	1.88
After the COE	4.38	2.32	5.87	0.19	0.23	6.71	1.47	2.31
$BA^{Treatment}$	1.30	0.57	1.30	0.07	0.08	2.49	0.60	0.43
Control group								
Before the COE	0.75	0.25	0.87	0.00	0.02	1.84	0.40	0.54
After the COE	1.34	0.80	1.85	0.04	0.07	3.26	0.66	1.12
$BA^{Control}$	0.59	0.55	0.97	0.03	0.05	1.42	0.26	0.58
The DID estimation results	0.709*** (0.126)	0.0221 (0.118)	0.325 (0.297)	0.0386* (0.0201)	0.0325 (0.0210)	1.065*** (0.179)	0.336*** (0.0827)	−0.150 (0.0987)

Robust standard errors are shown in parentheses

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 3 The DID estimation results for the citation per paper

	Adopted in FY 2002				Adopted in FY 2003			
	Life sciences	Information sciences	Chemistry, material sciences	Humanities	Social sciences	Medical sciences	Mechanical engineering	Mathematics and physics
Treatment group								
Before the COE	28.56	5.89	13.75	0.79	0.70	25.22	3.63	9.20
After the COE	37.39	8.12	16.19	0.96	0.99	31.17	5.43	10.59
$BA^{Treatment}$	8.83	2.23	2.44	0.17	0.29	5.94	1.80	1.39
Control group								
Before the COE	4.88	0.82	3.93	0.00	0.16	13.66	1.37	6.80
After the COE	9.04	1.56	6.26	0.00	0.33	16.17	2.75	10.65
$BA^{Control}$	4.15	0.74	2.32	0.00	0.17	2.51	1.38	3.85
The DID estimation results	4.674*** (1.456)	1.488** (0.716)	0.118 (0.888)	0.169 (0.193)	0.117 (0.201)	3.435** (1.368)	0.419 (0.443)	−2.457** (1.149)

Robust standard errors are shown in parentheses

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4 The increased effects classified according to the universities

	Life sciences	Information sciences	Chemistry, material sciences	Humanities	Social sciences	Medical sciences	Mechanical engineering	Mathematics and physics
The average increased effects in the number of papers								
National and public university	0.7 (22 teams)	0.1 (15 teams)	0.3 (18 teams)	0.1 (14 teams)	0.05 (17 teams)	1.1 (26 teams)	0.4 (17 teams)	−0.1 (21 teams)
Private university	0.1 (6 teams)	0.0 (5 teams)	0.4 (3 teams)	0.0 (6 teams)	−0.03 (8 teams)	0.7 (9 teams)	−0.1 (5 teams)	−0.3 (2 teams)
The average increased effects in the citations								
National and public university	6.4 (22 teams)	0.9 (15 teams)	−0.1 (18 teams)	0.2 (14 teams)	0.29 (17 teams)	4.1 (26 teams)	0.9 (17 teams)	−2.4 (21 teams)
Private university	1.5 (6 teams)	4.8 (5 teams)	1.0 (3 teams)	0.1 (6 teams)	−0.03 (8 teams)	2.6 (9 teams)	−1.0 (5 teams)	−2.0 (2 teams)

Table 5 The increased effects classified according to the major in the number of papers

	Classified according to the major						
Life sciences	0.7 (life sciences, 3 teams)	1.7 (biological sciences, 4 teams)	0.3 (biological mechanisms and functions, 5 teams)		0.4 (others, 11 teams)		
Information sciences	0.1 (school of information sciences, 7 teams)	0.0 (school of engineering, 6 teams)	−0.1 (school of science and engineering, 6 teams)				
Chemistry, material sciences	0.5 (school of science, 5 teams)	0.1 (school of engineering, 9 teams)	0.3 (school of science and engineering, 5 teams)				
Humanities	0.0 (literature and culture, 15 teams)	0.0 (education, 2 teams)	0.4 (human development, behavioral studies, 3 teams)				
Social sciences	−0.03 (law and politics, 8 teams)	0.09 (economics and management, 11 teams)	0.13 (economics, 8 teams)	−0.02 (management, 3 teams)	0.05 (policy studies, 5 teams)	−0.03 (sociology, 2 teams)	−0.15 (arts and letters, 1 team)
Medical sciences	1.5 (medical science, 22 teams)	−1.3 (nursing science, 3 teams)	−0.6 (medical and dental, 2 teams)		0.7 (others, 2 teams)		
Mechanical engineering	0.1 (environmental engineering, 9 teams)	0.2 (mechanical engineering, 8 teams)	0.5 (others, 2 teams)				
Mathematics and physics	−0.4 (mathematical sciences, 6 teams)	−0.3 (physics, 9 teams)	0.2 (geography, earth sciences, 8 teams)				

Table 6 The increased effects classified according to the major in the citations

	Classified according to the major						
Life sciences	8.8 (life sciences, 3 teams)	0.3 (biological sciences, 4 teams)	8.5 (biological mechanisms and functions, 5 teams)		7.8 (others, 11 teams)		
Information sciences	1.6 (school of information sciences, 7 teams)	0.3 (school of engineering, 6 teams)	4.2 (school of science and engineering, 6 teams)				
Chemistry, material sciences	0.8 (school of science, 5 teams)	−0.7 (school of engineering, 9 teams)	0.2 (school of science and engineering, 5 teams)				
Humanities	0.0 (literature and culture, 15 teams)	−0.6 (education, 2 teams)	1.3 (human development, behavioral studies, 3 teams)				
Social sciences	−0.05 (law and politics, 8 teams)	0.48 (economics and management, 11 teams)	0.30 (economics, 8 teams)	0.98 (management, 3 teams)	0.27 (policy studies, 5 teams)	−0.12 (sociology, 2 teams)	−0.91 (arts and letters, 1 team)
Medical sciences	3.7 (medical science, 22 teams)	−1.6 (nursing science, 3 teams)	5.4 (medical and dental, 2 teams)		−0.5 (others, 2 teams)		
Mechanical engineering	1.5 (environmental engineering, 9 teams)	−0.3 (mechanical engineering, 8 teams)	−0.1 (others, 2 teams)				
Mathematics and physics	−3.2 (Mathematical sciences, 6 teams)	−2.4 (Physics, 9 teams)	−1.7 (Geography, earth sciences, 8 teams)				

teams shown to have significant impact on the number of papers, in our estimation. Further, the two teams that had received the highest peer review assessment were among the five teams shown to have significant effects on the citation counts, in our estimation.

Information sciences and electrical and electronic engineering

In the field of information sciences, we found a statistically significant increase in the number of citations. The main results are given below.

- The number of papers per researcher per year increased from 1.75 to 2.32 after the introduction of the COE program (Δ0.57). The COE effect accounted for 0.02, which was not statistically significant. The citation counts per paper increased from 5.89 to 8.12 after the implementation of the program. Of the total increase (Δ2.23), the COE effect was 1.49, which was statistically significant at the 5 % level.
- We obtained statistically significant results for the number of papers by four teams out of a total of 20. Statistically significant results were also obtained for the citation counts of five teams.
- Within national universities (15 teams), the number of papers per researcher per year increased by 0.1; however, no such increase was found for the teams in the private universities (5 teams). Further, citation counts per paper increased by 0.9 and 4.8, respectively (see Table 4).
- Classification of effects according to graduate course showed that graduate schools of science and engineering witnessed the highest increase in citation counts (see Tables 5, 6).
- Comparison of our estimation results with the peer review analysis showed the following: one team, among four, that received the highest peer review assessment had a significant effect on the number of papers, in our model. Similarly, the team that received the highest peer review assessment from among five teams had a significant impact on the citation counts, in our analysis.

Chemistry and material sciences

In the fields of chemistry and material sciences, no statistically significant increase in numbers of papers or citation counts was observed. The main results are summarized below.

- The number of papers per researcher per year increased from 4.57 to 5.87 (Δ1.30). The COE effect was 0.32, which was not statistically significant. Citation counts per paper increased from 13.75 to 16.19 after the implementation of the program. COE was responsible for 0.12
- of the total increase (Δ2.44), which was not statistically significant.
- Statistically significant results were obtained for the number of papers authored by two teams out of a total of 21 teams. Results were also statistically significant for the citation counts of one team.
- Within national universities (18 teams) and private universities (3 teams), the number of papers per researcher per year increased by 0.3 and 0.4, respectively. The citation counts per paper decreased by 0.1 and increased by 1.0, respectively (see Table 4).
- Graduate schools of science experienced the highest increase both in the number of papers and citation count (see Tables 5, 6).

- The team that received the highest peer review assessment among two teams had significant effects, in our model, on the number of papers. Further, one team that did not receive the highest peer review assessment had significant effects, in our model, on the citation counts.

Humanities

Within humanities, we found a statistically significant increase in the number of papers. The main results are as follows.

- The number of papers per researcher per year increased from 0.12 to 0.19 after the introduction of COE program (Δ0.07). The COE effect was 0.04, statistically significant only at the 10 % level. Citation counts per paper increased from 0.79 to 0.96 (Δ0.17). Although COE effect was exclusively responsible for this increase in citations, the value was not statistically significant.
- We obtained statistically significant results for the number of papers, which were written by 2 out of 20 teams. However, no significant results were obtained for citation counts.
- Within national and public universities (14 teams), the number of papers per researcher per year increased by 0.1; however, no such increase was found for the teams in the private universities (6 teams). Similarly, citation counts per paper increased by 0.2 and 0.1, respectively (see Table 4).
- Among graduate courses, the field of psychology witnessed the highest increase both of in the number of papers and citation count (see Tables 5, 6).
- The team that received the highest peer review assessment among two teams had significant effect in terms of the number of papers, in our model.

Social sciences

In the field of social sciences, no statistically significant increase was observed for the number of papers or citation counts. The main results are given below.

- The number of papers per researcher per year increased from 0.15 to 0.23 after the COE program. Of the total increase (Δ0.08), the COE effect accounted for 0.03, which was not statistically significant. The citation counts per paper increased from 0.70 to 0.99 after the implementation of the program (Δ0.29). The COE effect was 0.12, which was not a statistically significant value.
- We obtained statistically significant results for the number of papers from 3 out of 25 teams. No such significant results were obtained for citation counts.
- In national universities (17 teams) and private universities (8 teams), the number of papers per researcher per year increased by 0.05 and decreased by 0.03, respectively. Citation counts per paper increased by 0.29 and decreased by 0.03, respectively (see Table 4).
- In terms of subject areas within the social sciences, economics and policy studies witnessed the highest increase in the number of papers and citations (see Tables 5, 6). The relatively higher academic performance in economics can be attributed to the fact that researchers in this field tend to publish in international English journals more often than researchers working in the other social sciences fields. The increase in the number of published economics papers was 0.12, statistically significant at the 1 % level.

Although the increase in the number of citation counts was 0.34, the value was not statistically significant.

- The three teams having significantly positive impact on the number of papers in our analysis had received the highest peer review assessments. However, eight other teams that did not have a significantly positive impact on the number of papers in our estimation fared well on the peer review assessment, because the training offered to the graduate students in their universities and the novelty of their research studies were highly evaluated.

Medical sciences

A statistically significant increase in the number of papers and citations was reported in the field of medical sciences. The main results are as under.

- The number of papers per researcher per year increased from 4.22 to 6.71 once the COE program was introduced (Δ2.49). The COE effect was 1.06, which was statistically significant at the 1 % level. Citation counts per paper increased from 25.22 to 31.17. The COE effect was 3.43 of the total increase (Δ5.94), which was statistically significant at the 5 % level.
- We obtained statistically significant results for the number of papers from 15 teams out of a total of 35 teams. Statistically significant results were also obtained for the citation counts by eight teams.
- In national and public universities (26 teams) and private universities (9 teams), the number of papers per researcher per year increased by 1.1 and 0.7, and citation counts per paper increased by 4.1 and 2.6, respectively (see Table 4).
- Graduate schools of medicine accounted for the highest increase in the number of papers and citations (see Tables 5, 6).
- Eleven of 15 teams that had a significantly positive effect on the number of papers in our model had received the highest peer review assessment. Similarly, five of eight teams with significantly positive effects on the citation counts, in our model, had received the highest the peer review assessment.

Mechanical, civil, architectural and other fields of engineering

In mechanical engineering and allied fields, we found a statistically significant increase in the number of papers. The following are the main results.

- The number of papers per researcher per year increased from 0.87 to 1.47 after the introduction of the COE program. Of the total increase (Δ0.60), the COE effect was 0.34, statistically significant at the 1 % level. Moreover, the citation counts per paper increased from 3.63 to 5.43 after the program was launched (Δ1.80). The COE effect was 0.42, which was not statistically significant.
- In terms of the number of papers, we obtained statistically significant results from 7 out of a total of 22 teams. For citation counts, we obtained statistically significant results from two teams.
- Comparison of the COE effects between national and public universities (17 teams) and private universities (5 teams) showed that the number of papers per researcher per year increased by 0.4 and decreased by 0.1, respectively. Further, citation counts per paper increased by 0.9 and decreased by 1.0, respectively (see Table 4).

- Out of a total of seven teams that had a significant impact on the number of the papers in our model, three teams had received a high peer review assessment. Similarly, the team that received the highest peer review assessment was among the two teams that had a significantly positive impact on the citation counts, in our model.

Mathematics, physics, and earth sciences

In the fields of mathematics, physics and earth sciences, no statistically significant increase was observed in the number of papers or citations. The main results are summarized below.

- The number of papers per researcher per year increased from 1.88 to 2.31 after the implementation of the COE program ($\Delta 0.43$). We found a negative COE effect of -0.15, which was not statistically significant. The citation counts per paper increased from 9.20 to 10.59 once the program began. The COE effect accounted for -2.46 of the total increase ($\Delta 1.39$). Because the before–after outcomes in the control group ($BA^{Control}$) exceeded those of the treatment group ($BA^{Treatment}$), these fields have negative effects (see Eq. 5).
- We obtained statistically significant results for the number of papers authored by 1 out of a total of 23 teams. No such significant results were obtained for the citation counts in any team.
- In national and public universities (21 teams) and private universities (2 teams), the number of papers per researcher per year decreased by 0.1 and 0.3, respectively. Further, the citation counts per paper decreased by 2.4 and 2.0, respectively (see Table 4).
- Comparison between our estimation results and those of the peer review showed that one team that had not received the highest peer review assessment had significant effects on the number of papers in our model. The negative evaluation in the peer review assessment was attributed to a decrease in the enrollment of doctoral course students and lack of novelty in research achievements, which only seemed to extend conventional research.

Discussion and conclusion

The main results of this paper are summarized as follows. In the fields of life sciences, humanities, medical sciences and mechanical engineering, we observed a statistically significant increase in the number of papers as a result of the COE program. A statistically significant increase in citation counts was also observed in the fields of life sciences, information sciences, and medical sciences.

The results above confirm that the positive impact of the 21st COE program, measured in terms of increase in the number of papers and citation counts, differs across research fields. The fields of life sciences and medical sciences have experienced the maximum impact of the COE program, both in terms of number of papers and citations. Having said that, one must also note that research cycles are remarkably different across fields, and the level and scale of the achievements in each research field before the introduction of the program was not the same (see Table 2). For instance, there are a great number of Japanese books and journals in the fields of social sciences and humanities.[6] In the field of social sciences and humanities, there are fewer publications in the control group, and thus, we should be cautious when interpreting the results. Haddow and Genoni (2010) indicated that the database coverage of journals in the social sciences was poor, whereas natural sciences journals were more easily cited, with shorter time lag between publication and citation.

One of the limitations of this study is that we may not have succeeded in accurately capturing the research trends specific to each field of study as we did not include the publications in proceedings. Only journal articles were used in this analysis, because they are considered as final achievements. We calculated the ratio of publications in proceedings for the teams that had a significantly positive impact on the papers and citations. In the field of life sciences, this ratio was 5 %, whereas it was 44 % in the information sciences. In addition, this analysis follows the field classification by the COE program, and various fields are mixed within a team. Thus, we note that in the fields of information sciences and electrical and electronic engineering, the number of papers and citations may differ within a field.

The differences, as mentioned above, in the research and education styles across fields make it difficult to determine the field that received the highest policy impact. One of the best approaches to adjust for the differences among fields and to compensate for this demerit of quantitative bibliometrics is the peer review. Bibliometric indexes play a supporting role and supplement the peer review with important information (Rinia et al. 1998). Thus, evaluation by a field-specific expert, as in a peer review, and econometric analysis, based on an individual researcher's achievements, have different merits and demerits.

As previously mentioned, there are various arguments about correlations between bibliometric indicators and peer reviews in previous studies we identified a few discrepancies between the results of our bibliometric analysis and the peer review in our study. For instance, despite their significantly positive impact on the number of papers or citations, some teams did not fare well in the peer review. This was because, peer reviews typically offered unfavorable evaluations for aspects related to the training of researchers and their collaborative research within a team. In other cases, the peer reviewers felt that research papers lacked novelty or original thought.

On the other hand, some teams that did not have a statistically significant impact on the number of papers or citation counts in our model received a favorable peer review assessment. This could be attributed to the highly positive evaluation of the training and research uniqueness.

Our study also has the following limitations. First, the educational component of the COE program was as important as the research component, and it aimed to develop excellent researchers. Although a peer review considers these aspects, we did not use any index to evaluate these educational aspects of the program. Second, since we used only two indexes, the number of papers and citations, it is difficult to evaluate the creativity and novelty of research activities through number of papers and citations in a short-term (5 years) evaluation. For a more effective evaluation, both peer reviews and econometric evaluation methods should be used together.

[6] In future, we may perform a DID estimation using the number of Japanese books in these fields along with English only journal papers and citations, from another database.

References

Anderson, R., Narin, F., & McAllister, P. (1978). Publication ratings versus peer ratings of universities. *Journal of the American Society for Information Science, 29*(2), 91–103.

Bozeman, B., & Gaughan, M. (2007). Impacts of grants and contracts on academic researchers' interactions with industry. *Research Policy, 36*(5), 694–707.

Cameron, A., & Trivedi, P. (2005). *Microeconometrics: Methods and applications* (pp. 860–898). New York: Cambridge University Press.

Committee for Research Evaluations. (2008). Present status of research evaluations and its future in Japan. Science Council of Japan. http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-20-t51-3e.pdf

Dawid, A. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society Series B (Methodological), 41*(1), 1–31.

Dietz, J., & Bozeman, B. (2005). Academic careers, patents, and productivity: industry experience as scientific and technical human capital. *Research Policy, 34*(3), 349–367.

Gaughan, M. (2009). Using the curriculum vitae for policy research: an evaluation of national institutes of health center and training support on career trajectories. *Research Evaluation, 18*(2), 117–124.

Gaughan, M., & Bozeman, B. (2002). Using curriculum vitae to compare some impacts of NSF research grants with research center funding. *Research Evaluation, 11*(1), 17–26.

Gaughan, M., & Ponomariov, B. (2008). Faculty publication productivity, collaboration, and grants velocity: using curricula vitae to compare center-affiliated and unaffiliated scientists. *Research Evaluation, 17*(2), 103–110.

Gibbons, M., & Georghiou, L. (1987). *Evaluation of research: A selection of current practices*. Paris: Organisation for Economic Co-operation and Development.

Haddow, G., & Genoni, P. (2010). Citation analysis and peer ranking of Australian social science journals. *Scientometrics, 85*(2), 471–487.

Hall, B., Jaffe, A., & Trajtenberg, M. (2000). Market value and patent citations: A first look. *NBER Working Paper No 7741*. http://www.nber.org/papers/w7741.pdf?new_window=1

Hall, B., Jaffe, A., & Trajtenberg, M. (2001). The NBER patent citations data file: Lessons, insights and methodological tools. *NBER Working Paper No 8498*. http://www.nber.org/papers/w8498.pdf?new_window=1

Harhoff, D., Narin, F., Scherer, F., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *The Review of Economics and Statistics, 81*(3), 511–515.

Heckman, J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions. In: J. Heckman & B. Singer (eds.), *Longitudinal analysis of labor market data* (pp. 156–245). New York: Cambridge University Press.

Jaffe, A., & Lerner, J. (2001). Reinviting public R&D: patent policy and the commercialization of national laboratory technologies. *RAND Journal of Economics, 32*(1), 167–198.

Jaffe, A., & Trajtenberg, M. (1996). Flows of knowledge from universities and federal laboratories: modeling the flow of patent citations over time and across institutional and geographic boundaries. *Proceedings of the National Academy of Sciences, 93*(23), 12671–12677.

Jaffe, A., Trajtenberg, M., & Fogarty, M. (2000). The Meaning of patent citations: Report of the NBER/case-western reserve survey of patentees. *NBER Working Paper No 7631*. http://www.nber.org/papers/w7631.pdf?new_window=1

Kostoff, R. (1994). Federal research impact assessment: state-of-the-art. *Journal of the American Society for Information Science, 45*(6), 428–440.

Lanjouw, J., & Schankerman, M. (1999). The Quality of ideas: Measuring innovation with multiple indicators. *NBER Working Paper No 7345*. http://www.nber.org/papers/w7345.pdf?new_window=1

Lee, M.-J. (2005). *Micro-econometrics for policy, program, and treatment effects*. New York: Oxford University Press.

Meho, L., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: web of science versus scopus and google scholar. *Journal of the American Society for Information Science and Technology, 58*(13), 2105–2125.

Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the web of science for coverage of the social sciences' literature. *Journal of Informetrics, 1*(2), 161–169.

Oppenheim, C. (1995). The correlation between citation counts and the 1992 research assessment exercise ratings for British library and information science university departments. *Journal of Documentation, 51*(1), 18–27.

Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for british research in genetics. Anatomy and archaeology. *Journal of Documentation, 53*(5), 477–487.

Opthof, T., & Leydesdorff, L. (2011). A comment to the paper by Waltman et al. Scientometrics 87 467–481, 2011. *Scientometrics, 88*(3), 1011–1016.

Rinia, E., van Leeuwen, Th, van Vuren, H., & van Raan, A. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria evaluation of condensed matter physics in the Netherlands. *Research Policy, 27*(1), 95–107.

Rinia, E., van Leeuwen, Th, van Vuren, H., & van Raan, A. (2001). Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. *Research Policy, 30*(3), 357–361.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Stock, J., & Watson, M. (2007). *Introduction to econometrics: International edition* (2nd edn., pp. 468–519). Boston: Pearson/Addison-Wesley.

Van Raan, A. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics, 67*(3), 491–502.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data* (pp. 603–643). Cambridge: The MIT Press.

Zhu, J., Meadows, A., & Mason, G. (1991). Citations and departmental research ratings. *Scientometrics, 21*(2), 171–179.