European Commission

JRC TECHNICAL REPORTS

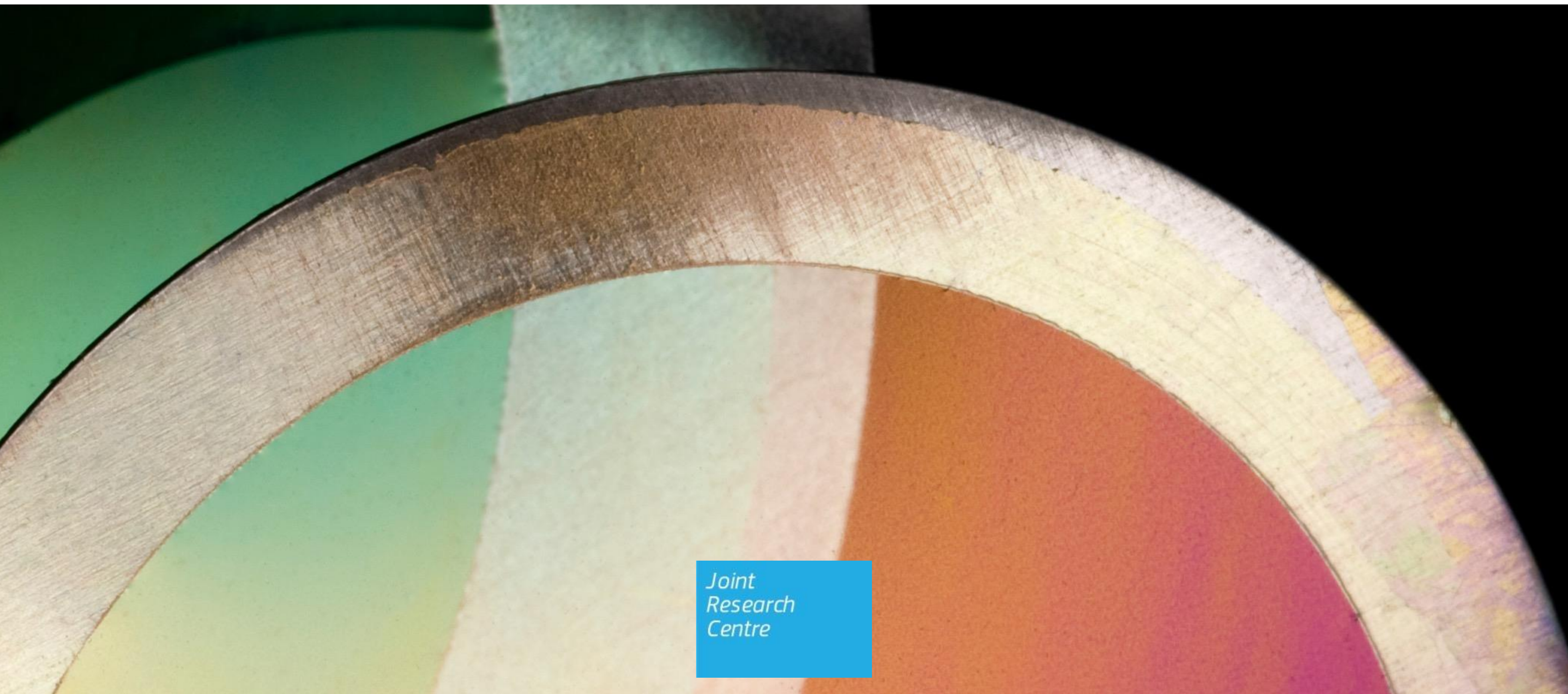
JRC Digital Economy Working Paper 2018-05

Does Online Access Promote Research in Developing Countries? Empirical Evidence from Article-Level Data

Frank Mueller-Langer[1,2], Marc Scheufen[3] and Patrick Waelbroeck[4,5]

[1] European Commission, Joint Research Center, Digital Economy Unit, Seville
[2] Max Planck Institute for Innovation and Competition, Munich
[3] Ruhr University Bochum
[4] Télécom ParisTech
[5] CESIfo, Munich

June 2018



Joint Research Centre

Contact information
European Commission, Joint Research Centre
Address: Edificio Expo. c/Inca Garcilaso, 3. 41092 Seville (Spain)
E-mail: frank.muller-langer@ec.europa.eu
Tel.: +34 9544-88731

JRC Science Hub
https://ec.europa.eu/jrc

JRC112230

Seville, Spain: European Commission, 2018

How to cite: Mueller-Langer, Frank, Marc Scheufen and Patrick Waelbroeck, Does Online Access Promote Research in Developing Countries? Empirical Evidence from Article-Level Data; Digital Economy Working Paper 2018-05; JRC Technical Reports.

Abstract

Universities in developing countries have rarely been able to subscribe to academic journals in the past. The "Online Access to Research in the Environment" initiative (OARE) provides institutions in developing countries with free online access to more than 5,700 environmental science journals. We analyze the effect of OARE on scientific output in five developing countries. We apply difference-in-difference-in-differences estimation using a balanced panel with 161,450 observations derived from 36,202 journal articles published by authors affiliated with 2,490 research institutions. Our approach allows us to explore effects across scientific fields, i.e. OARE vs. non-OARE fields, within institutions and before and after OARE registration. We benefit from the fact that variation in online access to scientific literature is exogenous at the level of scientific fields. Additional self-selection issues are dealt with by using an endogenous binary variable model estimated by a Bayesian Markov-Chain-Monte-Carlo method. We provide evidence for a positive marginal effect of online access via OARE on publication output that ranges between +48% and +57%. Our results suggest that the most productive institutions benefit the most from OARE while the least productive institutions barely benefit from it.

Keywords: Online access, scientific productivity, difference-in-difference-in-differences (DDD) estimation, Bayesian Markov-Chain-Monte-Carlo (MCMC) estimation

JEL codes: L17, O33

1. Introduction

While global online access has laid the groundwork for involving all nation-states in science, universities in developing countries have rarely been able to subscribe to academic journals in the past (Annan, 2004). For instance, most libraries in Sub-Saharan African countries had no access to any scientific journal for years (Suber and Arunachalam, 2005). The Online Access to Research in the Environment (OARE) initiative seeks to provide free or reduced-fee online access for researchers of registered institutions in the field of environmental science. It was launched by the United Nations Environment Programme (UNEP) and Yale University in October 2006. We focus our analysis on the five developing countries that are most active in terms of both publishing (number of articles in Web of Science) and registration with the OARE initiative: Kenya, Nigeria (Sub-Saharan Africa) and Bolivia, Ecuador, Peru (South America). We investigate the impact of the OARE initiative on local scientific output. All developing countries are eligible, but the initiative distinguishes between so-called Group A and Group B countries. Registered research institutions in Group A countries (gross national income (GNI) per capita below $1,600) receive free online access to all journals that are available under the OARE initiative whereas institutions in Group B countries (GNI per capita below $5,000) receive access for a reduced fee of $1,000 per year.

Using bibliometric article-level data from Web of Science (WoS) and OARE registration data from January 2000 to June 2012, we analyze the impact of OARE on the publication output of research institutions. Our identification strategy is based on the fact that OARE limits free or reduced-fee online access to environmental (OARE) journals. We use a difference-in-difference-in-differences (DDD) estimation that explores differences in publication output across OARE and non-OARE fields within institutions that registered with the OARE initiative and those that did not – before and after joining OARE. The underlying idea is that only researchers working on environmental issues can be impacted by free or reduced-fee online access to OARE journals after an institution has registered with OARE. In contrast, other scientific fields within the same institution (and OARE fields within the same institution but before OARE registration) will not benefit from the OARE program. Based on this underlying approach, we compare the scientific production of researchers in OARE fields in a given OARE member institution with the scientific production of researchers working on other fields at the same institution and researchers in non-member institutions before and after joining the initiative.

The DDD method has the advantage that it deals with concerns regarding self-selection at the institutional level that could imply that more productive universities might be more likely to register with OARE. We mitigate these concerns by looking at OARE and non-OARE disciplines in the same institution. Any remaining endogenous self-selection concerns are dealt with by using a Bayesian estimation method that explicitly models the correlation between unobserved variables, controlling for possible self-selection of institutions into the OARE initiative. This is a novel approach to endogenous self-selection in treatment programs.

We find that OARE membership increases the overall number of publications by a research institution by +48% to 57%. However, the most productive institutions benefit most from OARE membership while the least productive institutions barely benefit from it. These findings may have important policy implications as a higher productivity level in academic research may have a positive effect on economic growth and other means of economic prosperity, e.g. environmental innovation (Romer, 1986 & 1990, Griliches, 1957 & 1992; Jaffe, 1989).[1] The remainder of the paper is organized as follows: Section 2 relates our work to the literature on the economics of science, innovation and economic growth. In Section 3, we provide an overview of the data. Section 4 describes the methodology and the variables under study. In Section 5, we present the results of our empirical analysis and discuss robustness checks. Section 6 concludes.

2. Related literature

Access to scientific research has recently attracted widespread interest from economics scholars (Furman and Stern, 2011; McCabe and Snyder, 2015; Sorensen, 2004) and policy-makers (European Commission, 2012). In particular, open access (OA) has been subjected to a broad discussion on whether it is a promising new business model in the digital economy (Suber, 2012; Scheufen, 2015).[2]

---

[1] We will elaborate on the link between academic research and economic growth in section 2.

[2] Two arguments mainly drive this debate. First, with the advent of the Internet and the development of technologies to digitize information goods, scientific journal publishers have found new means to price discriminate ("big deals"), which has led to a sharp increase in journal subscription prices (Bergstrom and Bergstrom, 2004; Ramello, 2010) and hence higher costs of access to research. In contrast, OA provides free

The literature on open access can broadly be structured along three lines of research: studies investigating the effects of different publishing models (Shavell, 2010[3]; Jean and Rochet, 2010); studies analyzing the impact of different publishing models on readership and citations (Gaulé and Maystre, 2011; McCabe and Snyder, 2014 & 2015; Mueller-Langer and Watt, 2018); and studies directed towards a scientist's attitude and behavior regarding OA publishing (Hanauske et al., 2007; Eger et al., 2015). Our paper seeks to contribute to the first line of research. In particular, we study the effects of a change in the ability of researchers in developing countries to access academic works. We analyze the impact of this change before and after these researchers' institutions joined the OARE initiative, and we compare the results to those disciplines within institutions for which the access mode remained unchanged over time. Our research discusses the impact of free or reduced-fee online access on scientific production in developing countries, for which we find little prior literature. However, the need for such research is emphasized by Annan (2004). Our DDD approach allows us to examine the effect of OARE controlling for article characteristics and institutional characteristics such as rank, city population and the distance to the largest domestic city. Evans and Reimer (2009a) emphasize the need to further assess the role of open access in developing countries. Evans and Reimer (2009b, p. 5) show that "lower-middle-income countries tend to much more frequently cite freely available journals, but the poorest countries do not." Thus, scientists in the poorest countries seem to have virtually no access to online content. Evans and Reimer (2009a) suggest that poor infrastructure and slow internet access may explain this difference in citation rates. McCabe and Snyder (2015) criticize their paper, arguing that Evans and Reimer (2009a) do not control for citation trends. Our approach complements the two papers, as we analyze both input and output trends of access to academic

---

and unrestricted access to academic works (McCabe and Snyder, 2005 & 2014). Second, the copyright system that is behind these pricing schemes is built on the idea that commercial exclusivity granted by copyright generates the main incentive for the creator of a copyright work. Researchers, in contrast, are primarily motivated by reputation rather than by financial gains. Especially for journal articles, authors typically do not receive any royalties, since the copyright is generally transferred to the publisher (Shavell, 2010).

[3] Shavell (2010) argues that (a) readership is higher under open access, (b) a higher readership increases scholarly esteem, (c) research institutions would bear the costs of a shift towards the "author pays" model and (d) there are several reasons why legal action is necessary to facilitate a change towards an universal OA regime. Several researchers have critically assessed the assumptions made in Shavell (2010). See Mueller-Langer and Scheufen (2013) for a review.

works for researchers in the developing world.[4] We contribute to this strand of literature by investigating the role of free and reduced-fee online access in developing countries on scientific production.

Our paper also contributes to the literature in the broader field of economics of science and innovation investigating the role of science and scientific research in the advancement of technologies and hence in fostering economic growth (Dasgupta and David, 1994; Dosi, 1988; Merton, 1973; Murray et al., 2016).[5] In general, Romer (1986, 1990) highlights the role of academic research as a major factor for technological innovations and hence for economic growth. Before Romer, the literature especially by Solow (1956) and Swan (1956) were able to explain the role of academic research for economic growth by means of a residual as growth was exogenously determined. Romer's endogenous growth theory endogenizes technological progress by emphasizing the relevance of spill-overs from academia. Accordingly, when free online access increases scientific output, this eventually may have a positive effect on innovation and economic growth. Extending on Romer (1990) several authors have emphasized the importance of knowledge spillovers from science for economic growth (Griliches, 1992; Jaffe, 1989; Audretsch and Feldman, 1996; Acs et al., 1994).[6] However, these spillovers cannot be taken for granted as we find that only 5 percent of eligible institutions are OARE members.

3. Data and proceedings

3.1. The OARE initiative

The OARE initiative is led by the United Nations Environment Programme (UNEP) in partnership with major publishers in environmental science.[7] OARE was launched in October 2006. Today, OARE offers access to more than 5,710 peer-reviewed scientific journal titles published by 461 OARE partners in more than 100 eligible countries. In this regard, eligibility distinguishes between

[4] Input is measured by the relative number of cited OARE articles in a given article, while output is measured by the total number of articles of a given institution.

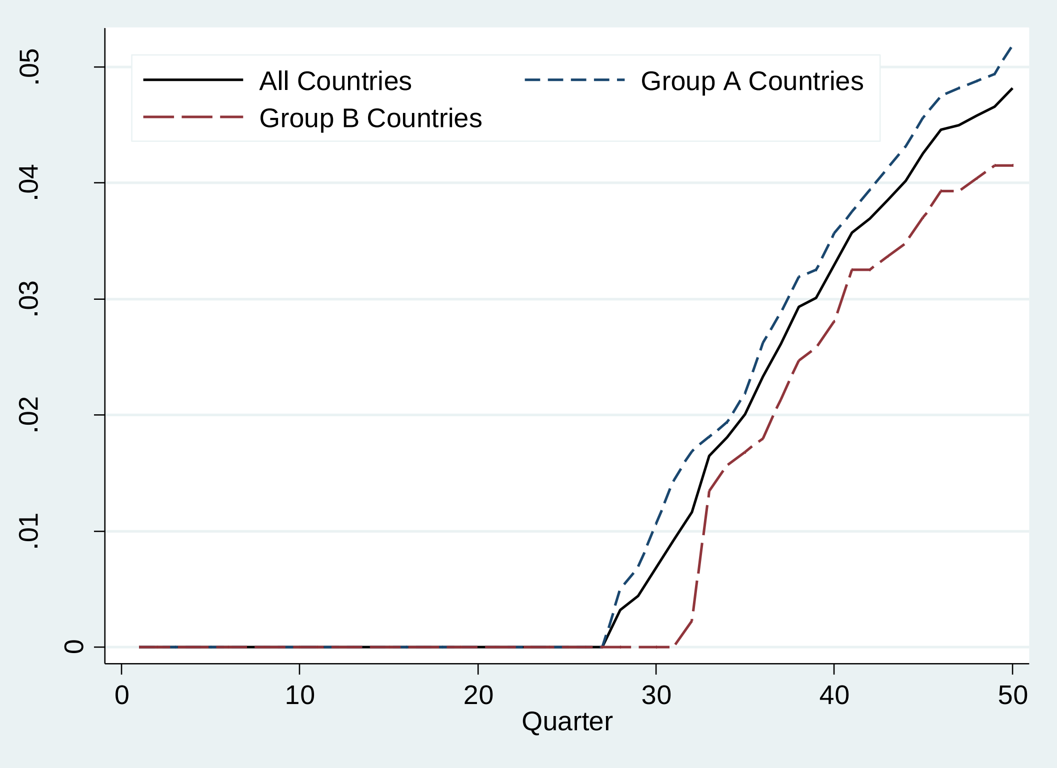
[5] See also Stephan (1996) for an overview of the economics of science literature.

[6] See Diamond (1994) for an overview on Zvi Griliche's contributions for understanding the economics of technology and growth. See also Geroski (2000), Hall (2004), Hall and Kahn (2003) and Mansfield (1961 & 1963).

[7] See http://oare.research4life.org/content/en/partners.php for an overview on the major partners of the OARE initiative.

Group A (free online access) and Group B (low-cost access) countries,[8] depending on the countries GNI per capita. Institutions in countries with a GNI per capita at or below $ 1,600 receive free access to the full content of journal articles, while institutions in countries with a GNI per capita below $ 5,000 pay a fee of $ 1,000 per year. However, institutions have to register to OARE in order to receive access. In creating awareness and knowledge, OARE offers additional training by means of courses and workshops for librarians and researchers for instance.

FIGURE 1 | OARE ADOPTION OVERALL AND BY COUNTRY GROUP



Adoption patterns of OARE (starting quarter: 28) for all countries (as given by the middle solid line) and by country group (dashed lines). Group A countries: Kenya, Nigeria and Bolivia. Group B countries: Ecuador and Peru. Figure 1 suggests that different institutions registered with OARE at different points in time. This implies that we have multiple cut-offs for before and after OARE depending on the respective institution. Figure 1 also implies that only about 5% of all eligible institutions joined OARE after a period of 5 ½ years. This points to the unused potential of the OARE initiative.

Figure 1 illustrates the rate of adoption of OARE over time (quarters) in all countries (as given by the middle, black solid line) and separately in Group A countries (upper blue dashed line) and

[8] Please note that countries can also convert from one group to the other if the GNI per capita changes over time. As such, Bolivia changed from group A to group B in 2017. For the time horizon under study, however, we do not find any group transitions.

Group B countries (lower red dashed line).[9] The rate of adoption is measured by the cumulative number of institutions that joined OARE in a given quarter divided by the total number of institutions, i.e. 2,490 institutions in all countries (Group A countries: 1,599; Group B: 891). Finally, it is worth noting that about 5% of all eligible research institutions in Group A countries and about 4% of all eligible research institutions in Group B countries had joined OARE in the last quarter under study (June 2012).[10]

3.2. Data

Our dataset is built from three main sources. First, we collected bibliometric article-level data from WoS for the five countries under study. We focus our analysis on five countries for the following reasons. On the one hand, we choose the most productive countries in terms of the total number of research articles from January 2000 to June 2012 for both geographical regions (Sub-Saharan Africa and South America). On the other hand, we look at countries that exceed a threshold of at least 20 OARE institutions in order to have variation across institutions within countries.[11] Second, we gathered institutional data including institutions' registration with OARE. Third, we extracted the rank of the institutions from the Ranking Web of Universities.

Regarding the first data source, we collected a panel dataset containing metadata for 36,202 research articles. The period under study starts in January 2000 (quarter 1) and ends in June 2012 (quarter 50). We obtain article metadata from WoS. The WoS data contain the institutions of the authors, the title of the paper, journal information (publication date, number of pages, volume number, issue number) and the number of citations. Overall, we have 2,490 institutions that published at least one article over the period under study.

We use article-level data for assigning different characteristics to each single article, accounting for the field of research, institutional affiliations of the authors, cooperation with authors from outside the developing world and other controls such as number of references, pages etc. Since the OARE

---

[9] We used Internet Archive's Wayback Machine to explore possible group changes over time. All countries under study remained in the same group for the period under study, i.e. 2000 to 2012.

[10] Note that the total number of eligible institutions refers to institutions that have observable research output in the period under study. We exclude non-research institutions from our analysis, i.e., we drop institutions that did not publish any journal article during the period under study.

[11] Please also note that the data creation process involved manual matching of institutions using different versions of search terms in Stata string matching functions. We will further elaborate on the data creation process in section 3.1.

initiative offers free or reduced-fee online access to research in environmental science, we create a dummy variable indicating whether an article falls under an OARE research area. We define an article as falling under an OARE research area if its "Research Area" provided by WoS also appears frequently in the titles of OARE journals. We proceed as follows. First, for all articles under study, we extract all terms from the WoS "Research Area" field. Second, we order these research area terms by frequency, i.e., we count how many articles in the data fall under a given single-word term (henceforth, WoS research area terms). For instance, in the case of articles of authors affiliated with Nigerian universities, the term "environmental" appears 2,179 times, whereas the term "architecture" appears once. Next, we extract the 200 most frequent terms that appear in the complete list of titles of OARE journals (henceforth, top 200 title terms). Matching these two lists (WoS research area terms and top 200 title terms), we obtain the top 50 OARE research areas. The top 50 OARE research areas are given by the 50 most frequent WoS research area terms that are also included in the top 200 title terms. Finally, we use a "top 50 OARE research"-dummy (which is one if an article falls under the top 50 OARE research fields, 0 otherwise). Distinguishing between OARE fields and non-OARE fields within a given institution allows us to explore the effect of online access to OARE journals on scientific output in OARE fields before and after OARE registration as well as in OARE fields as compared to non-OARE fields before and after OARE registration.[12]

Our sample contains all articles of researchers of the countries under study, including both single and multiple authored articles. However, dividing the share of a publication between different (local) authors to determine the respective contributions of authors is a challenging task for at least two reasons. First, there is no consensus within and across disciplines on how to account for multiple authorships. In particular, taking each author of a paper fully into account would overestimate the output produced. Creating a weight for multiple authored papers by dividing each publication by the number of authors, however, would also necessarily involve assumptions on the habits of co-authorship. In some disciplines (or publishing cultures), the order of authors has clear implications. Sometimes the first author or the last author is perceived as the "main author" of a

[12] See section 4.1 on the methodology and definition of our treatment variable.

research article. Other disciplines choose the order of authors alphabetically or by status. All of this makes it hard to operationalize multiple-authored papers from one country. Second, to the best of our knowledge, McCabe and Snyder (2015) is the only reference that explicitly deals with the issue of single versus multiple authors with respect to online access. The authors restrict their sample to single authors (from a local country) only, due to the difficulties in dividing the share of a multiple authored paper between the authors. However, to consider multiple co-authored articles in addition to single authored articles has two main advantages. First, only looking at single authored articles would substantially reduce our sample by 18,955 articles, that is, more than 50% of our dataset. Second, multiple co-authored articles may have different characteristics than single authored articles.

In the light of these advantages, we argue that the above-mentioned concerns can be overcome as follows. We account for multiple authorship by simply dividing the institutional share of each paper by the number of authors. For instance, a paper with two authors from two institutions leads to an increase in output of 0.5 for each of these institutions.[13] For robustness, we also provide the regression results for single authored papers in section 5.3 (Table 4). The results are remarkably similar.

To construct the (balanced) panel, we gather article level information by institution, field (OARE vs. non-OARE) and quarter for each country under study. For each country, we then merge rank and city information – including population and distance data – from separate datasets. Subsequently, we merge all individual country data into one dataset.[14] We distinguish country-specific information by generating a unique country ID for all countries. In a final step, we drop institutions that published during only one quarter. In total, we obtain 161,450 institution-discipline-quarter pairs, which constitute our unit of observation.

In assigning institutions to authors of articles from the countries under study, we use Stata string-matching functions, searching for snippets of institution names and abbreviations. In particular, we

[13] Note that due to the complexity of the manual string matching process we restrict our calculations to include multiple authored papers with up to 11 authors. This restriction does not reduce our sample by too much because 94% of all papers have 11 or less authors.

[14] We take the mean for the continuous variables, the max for the binary variables and the sum for the publication variable in performing the collapse command.

manually account for different versions and spellings of institutions, as WoS does not provide with a unique number or code to unambiguously identify a particular institution of interest. Most importantly, also spelling errors, case sensitivity as well as abbreviations impede an automatic matching of articles and institutions. Last but not least, we repeat the string matching process for each country file for each author level, accounting for up to 11 levels (11 authors for each article) and distinguishing different author level ids. We unambiguously identify 459 research institutions that are part of the Ranking Web of World Universities and/or OARE member institutions, forming the core universities for the string-matching process.[15] For each country under study, we find a large number of institutions that are neither included in the Ranking Web of Universities list nor in the list of OARE institutions. For these institutions, we generate unique institution IDs as follows. First, we order the institutions in a given country alphabetically. Second, we identify all instances of a given institution in the raw data. For instance, a given institution can have multiple versions because of abbreviations, use of different languages, or typos. Thereby, we also use the city where an institution is located to identify different versions of a given institution, manually assigning identical institution IDs in such cases.

Moreover, we assign institution IDs to track the relative position of an institution in the university ranking list. For a given country, a lower institution ID reflects a better rank. The rank variable, in addition, reflects the absolute worldwide position of the institution in the Ranking Web of World Universities. This ranking provides information on the performance of 22,123 research institutions worldwide on the basis of the web presence as well as the impact of institutions. The former aspect is particularly noteworthy, as web presence provides also a proxy for the technical expertise needed to set up online access to journals.

Finally, we assign city IDs to construct distance and population variables. To give an example, we identify 74 cities in Nigeria with a population of more than 100,000 inhabitants (pop variable) using the World Population Review (2017), listing population numbers for each city in each country of the world. In addition, we identify 64 cities from our Nigeria sample with fewer than 100,000 inhabitants. We assign city IDs 1 to 138 to the Nigerian cities, where a lower number denotes a

[15] In total, 163 institutions in Nigeria, 96 in Peru, 82 in Kenya, 62 in Ecuador and 56 in Bolivia.

larger population. As a further control, the variable *distance_1* was created by using Google maps and by computing the distance in km from the city in which an institution is located to the largest domestic city, as suggested by the first itinerary option by car. In addition, we create a *distance_2* variable indicating the distance of an institution's location to the next domestic million city.

4. Methodology and Variables

4.1. Methodology

We use a DDD approach and two different estimation methods in our analysis. First, we estimate the OARE effect using OLS regression analysis. Second, we model the OARE effect as an endogenous binary variable in a Bayesian Markov-Chain-Monte-Carlo (MCMC) simulation framework to account for potential self-selection into the OARE initiative.

4.1.1. DDD using OLS regression

In order to analyze the effect of the OARE initiative, we use a DDD method for comparing the change in research output for research fields in the treatment group (i.e. environmental sciences in registered institutions after OARE registration) with the change in research output for scientific fields in the control group (i.e. environmental sciences in registered institutions before OARE registration, non-environmental sciences in registered institutions and all research fields in unregistered institutions) before and after a given institution has registered (or not) with OARE. The intuition behind the DDD approach is the following. Within an OARE institution, only researchers working on environmental issues can be impacted by free or reduced-fee online access to environmental (OARE) journals after the institution has registered with OARE. In contrast, other scientific fields within the same institution (and OARE fields before OARE registration) will not benefit from the OARE program. Exploring effects of online access across scientific fields within a given institution mitigates concerns of self-selection at the institutional level, for instance, because better/more productive institutions might be more likely to register with OARE.

The dependent variable, $y_{s,t,r}$ is the log of the number of published articles by researchers from institution *s* in quarter *t* in research area *r*. We use the specification outlined in equation (1):

(1) $y_{s,t,r} = b_0 + b_{1,s,r}\, fe_{s,r} + \Sigma_t\, b_{2t}\, q_{s,t} + b_3\, treated_{s,t,r} + \Sigma_k\, b_{4,k}\, x_{k,s,t,r} + e_{s,t,r}$

with $k = 1,\ldots, K$; $t = 1, \ldots T$

where $fe_{s,r}$ are institution-discipline fixed effects. Quarter dummies are given by $q_{s,t}$ and T = 50. Variable $treated_{s,t,r}$ is our main variable of interest. It accounts for the fact that institutions registered with the OARE initiatives at different points in time and that other disciplines than environmental sciences in a given institution will not benefit from OARE. In other words, *treated* is 1 if an institution is an OARE institution and if articles of affiliated researchers are published in the OARE research area in a quarter after the institution registered with OARE (and 0 otherwise). $X_{k,s,t,r}$ are k control variables (k=1, …, K).[16] $e_{s,t,r}$ are unobservable effects assumed independent across s, t and r. When we refrain from including institution-discipline fixed effects, we include institutional and city characteristics such as worldwide rank, distance to largest domestic city/next 1-million city and city population.

4.1.2. DDD using Bayesian estimation to account for self-selection

We estimate the OARE effect using Bayesian estimation techniques based on a data augmentation MCMC algorithm described in Appendix 1. There are two equations. The first equation determines self-selection in the OARE initiative using a latent variable framework. The second equation is identical to equation (1). We assume that the unobserved variables of both equations follow a bivariate normal distribution with correlation coefficient r. The MCMC algorithm simulates the latent variable of the first equation to generate the endogenous binary treatment effect. The Bayesian approach explicitly deals with the correlation between the unobserved variables of the two equations. If there are any unobserved variables that determine whether an institution self-selects into the OARE program, the Bayesian method accounts for its potential endogeneity on the estimation of the treatment effect. This comes at the cost of increasing computing time since the self-selection and productivity equations are estimated at the same time.[17]

4.2. Definition of variables

Table 1 provides an overview of the variables under study and summary statistics at the institution-discipline-quarter level. Appendix 2 provides summary statistics by country group.

---

[16] For instance, we include article characteristics such as the number of pages, references, co-authors USA, co-authors EUR and OARE references (i.e. as an input variable).

[17] A typical MCMC algorithm run with 10,000 observations and 5,000 simulation iterations took 12-16 hours on an i5 2500K equipped computer.

Variables can be grouped into six categories: dependent variable, countries, main variable of interest, article characteristics, institutional characteristics and city characteristics.

TABLE 1 | SUMMARY STATISTICS

VARIABLES	mean	sd	min	max	N
Dependent variable					
# publications	0.216	1.477	0	93.82	161,450
Countries					
Kenya	0.261	0.439	0	1	161,450
Nigeria	0.268	0.443	0	1	161,450
Bolivia	0.122	0.327	0	1	161,450
Ecuador	0.123	0.328	0	1	161,450
Peru	0.227	0.419	0	1	161,450
Main variable of interest					
OARE treated (DDD)	0.010	0.100	0	1	161,450
Article characteristics					
# co-authors USA	0.728	1.593	0	37.25	161,450
# co-authors EUR	0.837	1.990	0	57	161,450
# OARE references	6.621	8.064	0	135	161,450
# pages	9.569	5.397	1	120	161,450
# references	31.76	18.51	0	293	161,450
Institutional characteristics					
Rank1: rank<=5,000	0.028	0.165	0	1	161,450
Rank2: 5,000<rank<=10,000	0.019	0.137	0	1	161,450
Rank3: 10,000<rank<=15,000	0.021	0.145	0	1	161,450
Rank4: 15,000<rank<=25,000	0.032	0.177	0	1	161,450
Rank5: rank=>25,000	0.899	0.302	0	1	161,450
rank, in 1,000	7.211	4.986	0.749	21.79	161,450
City characteristics					
Distance from largest domestic city, in 100 km	3.254	3.624	0	20.64	161,450
Distance from closest domestic city with more than 1 million inhabitants, in 100 km	1.967	3.189	0	20.64	161,450
Pop0: pop<=100, in 1,000	0.229	0.420	0	1	161,450
Pop1: 100<pop<=500, in 1,000	0.108	0.310	0	1	161,450
Pop2: 500<pop<=1,000, in 1,000	0.159	0.365	0	1	161,450
Pop3: 1,000<pop<=5,000, in 1,000	0.334	0.472	0	1	161,450
Pop4: pop>5,000, in 1,000	0.361	0.480	0	1	161,450

Data is aggregated at the institution-discipline-quarter level. The institution-discipline-quarter pairs constitute the unit of observation. We take into account journal articles by both single and multiple local authors in the five countries under study.

4.2.1. Dependent variable

Our dependent variable, $y_{s,t,r}$, indicates the number of publications of institution *s* in quarter *t* in discipline *r*. In the regression, we take the log to avoid problems of heteroskedasticity.[18]

4.2.2. Independent variables

Countries: We study 2,490 institutions from five countries of which two are located in Sub-Saharan Africa (Kenya and Nigeria) and three in South America (Bolivia, Ecuador, Peru). At the institution-discipline-quarter level, 52.9 % of our observations are from Sub-Saharan Africa.

Main variable of interest: treated is our main variable of interest. We construct this treatment variable by interacting three dummy variables. First, *OARE* indicates whether papers are written by authors affiliated with OARE institutions. We generate the OARE dummy by using the institution IDs of all institutions that are part of UNEP's list of OARE institutions. *OARE* (not reported in the table) takes on the value 1 if the respective institution of an article under study is an OARE institution and the value 0 otherwise. Second, the *after* dummy (not reported in Table 1) accounts for the registration date (in quarters) of a certain OARE institution. Its value is 1 if the article under study was written by an author affiliated with an OARE institution after the institution joined the OARE program and 0 otherwise. Third, we generate an OARE research field dummy capturing whether a particular article is within the top-50 OARE research areas or not. This allows us to compare differences within institutions, i.e. differences between disciplines that are core OARE research fields (e.g. environmental science) versus non-OARE fields of research (e.g. economics).

Article characteristics: #co-authors USA (*#co-authors EU*) indicates the average number of co-authors from the US (EU). Finally, *#OARE references* indicates the average number of references from OARE journals. That is, we consider references from OARE journals as an input variable. *#pages* indicates the average number of pages. The average number of references is indicated by *#references*.

Institutional characteristics: Five variables indicate the rank of an institution constructed from the Ranking Web of Universities (2014). *Rank1* represents the best institutions (rank≤5,000) whereas *Rank4* corresponds to institutions with the lowest reported ranks (15,000<rank≤25,000). *Rank5*

[18] The histogram of the log number of publications at the institution-discipline-quarter level is shown in Appendix 3.

indicates that an institution is not listed, which implies that its rank is above 25,000; these institutions are the least productive in scientific output.[19]

City characteristics: We use two different distance measures, indicating (1) the distance in 100 km of a given city to the largest domestic city (henceforth also *distance_1*) or (2) the distance in 100 km of a given city to the next domestic 1 million city (henceforth also *distance_2*).[20] City population dummies indicate the number of inhabitants of the city where an institution is located: *Pop0* indicates cities with less than 100,000 inhabitants whereas *Pop4* indicates cities with more than 5,000,000 inhabitants.

5. Empirical Analysis

5.1. The OARE Effect

We estimate the impact of OARE membership on scientific output by using eight different specifications in Table 2. Specifications (1) to (7) use OLS estimation,[21] whereas we apply the Bayesian MCMC method in column (8). Column (1) reports the OLS regression coefficients for the basic model, including the treatment variable as well as country and quarter dummy variables. We add article characteristics in (2), institutional rank information in (3), city population in (4), distance to the largest domestic city in (5) and distance to the next domestic city with more than 1 million inhabitants in (6). In specification (7), we include institution-discipline fixed-effects instead of country dummy variables and institutional and city characteristics (rank, population and distance). The last column (8) of Table 2 reports the coefficients estimated using the Bayesian MCMC algorithm using institution-discipline fixed effects. It can be directly compared with Specification (7).[22]

---

[19] Using rank categories instead of the actual rank has the advantage of being invariant to small variations in rank over time.

[20] We do not have distance information for 206 institutions, as the respective cities do not appear in Google maps. For these cities, we proxy the distance to the largest domestic city by taking the average distance in the respective country. We use the same approach for our alternative distance variable.

[21] We use the *xtreg* command in STATA. The institution-discipline-quarter pairs constitute the unit of observation.

[22] The MCMC algorithm was "warmed up" with 5,000 iterations and the next 5,000 iterations were used to compute the coefficients reported in Table 2.

TABLE 2 | OARE EFFECT

Model:	(1) Base	(2) + Article info	(3) +Rank	(4) +Popul.	(5) +Dist._1	(6) +Dist._2	(7) +Inst.-Disc. FE	(8) MCMC
Dependent variable:	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)	Log(w+1)
OARE treated (DDD)	0.403***	0.401***	0.397***	0.397***	0.397***	0.397***	0.389***	0.4505***
	(0.0350)	(0.0348)	(0.0346)	(0.0346)	(0.0346)	(0.0346)	(0.0341)	(0.0150)
# pages		0.000151	0.000176	0.000168	0.000167	0.000169	0.000788	0.000004
		(0.000731)	(0.000730)	(0.000731)	(0.000731)	(0.000731)	(0.000934)	(0.0002)
# references		-4.65e-05	-5.61e-05	-6.66e-05	-7.02e-05	-6.84e-05	-1.24e-05	-0.00009
		(0.000262)	(0.000261)	(0.000259)	(0.000259)	(0.000259)	(0.000323)	(0.00007)
# co-authors USA		0.00131	0.00140	0.00139	0.00136	0.00136	0.000343	0.0019**
		(0.00329)	(0.00311)	(0.00312)	(0.00312)	(0.00312)	(0.00314)	(0.0007)
# co-authors EUR		0.00667*	0.00660*	0.00656*	0.00657*	0.00656*	0.00807*	0.0084***
		(0.00381)	(0.00387)	(0.00387)	(0.00387)	(0.00387)	(0.00450)	(0.0008)
# OARE references		0.000576	0.000564	0.000622	0.000643	0.000636	0.000366	0.0001
		(0.000565)	(0.000566)	(0.000560)	(0.000560)	(0.000561)	(0.000699)	(0.0001)
Rank2: 5,000<rank<=10,000			-0.270***	-0.255**	-0.254**	-0.253**		
			(0.0997)	(0.0994)	(0.0993)	(0.0998)		
Rank3: 10,000<rank<=15,000			-0.296***	-0.278***	-0.279***	-0.278***		
			(0.0877)	(0.0871)	(0.0870)	(0.0871)		
Rank4: 15,000<rank<=25,000			-0.0749	-0.0622	-0.0613	-0.0606		
			(0.111)	(0.110)	(0.110)	(0.110)		
Rank5: rank=>25,000			-0.364***	-0.368***	-0.367***	-0.367***		
			(0.0775)	(0.0778)	(0.0777)	(0.0778)		
Pop1: 100<pop<=500, in 1,000				0.0369	0.0394	0.0393		
				(0.0298)	(0.0304)	(0.0307)		
Pop2: 500<pop<=1,000, in 1,000				0.0794**	0.0807**	0.0800**		
				(0.0365)	(0.0369)	(0.0368)		
Pop3: 1,000<pop<=5,000, in 1,000				0.0691**	0.0612**	0.0623**		
				(0.0316)	(0.0303)	(0.0309)		
Pop4: pop>5,000, in 1,000				0.0736**	0.0647**	0.0680**		
				(0.0295)	(0.0290)	(0.0292)		
Distance from largest domestic city, in 100 km					-0.00236			
					(0.00176)			
Distance from closest domestic city with > 1 million inh., in 100						-0.00189		
						(0.00190)		
Constant	0.127***	0.122***	0.446***	0.377***	0.393***	0.382***	0.0682***	0.0196***
	(0.0158)	(0.0162)	(0.0766)	(0.0802)	(0.0804)	(0.0800)	(0.00921)	(0.0066)
ρ								-0.260***
								(0.0567)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO	NO
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES	YES
Observations	161,450	161,450	161,450	161,450	161,450	161,450	161,450	16,150
R-squared	0.091	0.088	0.135	0.140	0.138	0.138	0.0756	-
Number of Inst_Discipline	3,229	3,229	3,229	3,229	3,229	3,229	3,229	323
Number of Inst	2,490	2,490	2,490	2,490	2,490	2,490	2,490	650

We use a balanced panel and take into account journal articles by both single and multiple local authors. Results on the impact of OARE membership (*treated*) on publication output of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru) from OLS DDD and Bayesian MCMC estimation methods. In (1) to (7), we use the *xtreg* command in STATA. The institution-discipline-quarter pairs constitute the unit of observation. Period under study: 1^{st} quarter 2000 to 2^{nd} quarter 2012. We obtain the findings on the OARE effect mentioned in the text by calculating the exponential of the *treated* coefficient minus 1. Reference country is Nigeria. Reference quarter is 36. Reference rank is *rank≤5000*. Reference population is *pop≤100*. Robust standard errors clustered at the institutional level (OLS) and standard errors of the marginal posterior distributions (Bayesian) reported in parentheses. Note that serial correlation is not an issue in our balanced panel because the large number of periods with 0 publication breaks any time correlation for any given institution. To make the MCMC procedure tractable, we ran the program over a random sample of 650 institutions. Other samples yielded similar results that are available upon request. Bayesian estimation is based on the estimation of two equations and does not produce as such a value for R-squared. However, since the standard deviation of the unobservable variable of the observation equation is almost identical to the value obtained in (7), the R-squared would be almost the same.

**p* < 0.1, ***p* < 0.05, ****p* < 0.01.

We find a positive and robust OARE effect that is statistically significant at the 1% level across all specifications.[23] The marginal OARE effect ranges from +48% in specification (7) to +57% in specification (8) where we use the MCMC method described in Appendix 1 that explicitly deals with self-selection into the OARE initiative.[24] Notably, the MCMC coefficient for *treated* (0.45) is larger than the coefficient in the base OLS specification (0.40). This is due to the statistically significant negative correlation of -0.26 between the unobserved variables of the self-selection equation and those of the publication output equation.[25] Unobserved variables in the self-selection equation include the hidden (administrative and informational) costs of joining the initiative, while unobserved variables in the main equation include hidden productivity factors. A negative correlation between the unobserved variables corresponds to a negative correlation between the hidden costs of joining the OARE initiative and the unobserved productivity variables at the institution level. We also ran the regressions separately for Group A and Group B countries (Appendix 5) and for each of the five countries (Appendix 6). The OARE treatment effect is positive and statistically significant for these subgroups. It is higher for institutions in Group A countries (i.e. free access countries) than for institutions in group B countries (i.e. reduced-fee countries).

Moving from column (1) to column (2), we consider the effects of article characteristics on publication output. Interestingly, cooperations with researchers from the EU have a positive and statistically significant effect (at the 10 percent level and smaller) on the publication output of institutions in developing countries.[26] In addition, this effect appears to be much smaller than the OARE effect.

R-squared remains almost the same (0.091 versus 0.088) and the OARE effect remains almost the same when we include article characteristics in (2). In contrast, R-squared increases almost by a

[23] All country dummy variables are negative, as the reference country Nigeria has the largest publication output.

[24] We obtain these results by calculating the exponential of the *treated* coefficient minus 1.

[25] We also find a statistically significant OARE effect of similar size when we use an unbalanced panel instead of a balanced panel (Appendix 4). In the balanced panel quarter-institution-discipline observations without any publications are treated as zero-publication observations. In contrast, these are treated as missings in the unbalanced panel. Thus, our results are robust to the inclusion of zero-publication observations.

[26] Looking at Appendix 6, we see that the EU-cooperation effect is mainly driven by Ecuador and Peru. A possible explanation for this result are cultural and economic ties of the two countries with Spain.

factor of two (from 0.088 to 0.135) while the OARE effect decreases only slightly from 0.401 to 0.397 when we add institutional rank information in (3).[27]

We also find that lower-ranked institutions are less productive in terms of publication output, since the coefficients associated with lower ranks (5000<*rank*≤10000, 10000<*rank*≤15000 and *rank >= 25,000*) as compared to the best rank category *rank* ≤5000 (reference category) are negative and statistically significant at the 1% level. The coefficient for 15000<*rank*≤25000 is also negative across specifications but not statistically significant.

In addition, we find a positive and statistically significant effect at the 5% level for institutions located in cities with more than 500,000 inhabitants, i.e. *Pop2*, *Pop3* and *Pop4*. This suggests that institutions located in larger cities in terms of population publish more research. The distance to the largest domestic city has a negative but not significant impact on output. Finally, the results reported in column (8) provide additional empirical support that collaborations with authors from Europe have a positive effect on scientific output in developing countries.[28]

5.2. Productivity of Institutions and the OARE Effect

We also ran the regressions separately for institutions with different levels of productivity. Results are reported in Table 3. Specifications (1), (2) and (3) report results for institutions with publications in 25 or fewer quarters. Specifications (4), (5) and (6) report results for institutions with publications in more than 25 quarters. Specifications (5), (6) and (7) of Table 2 serve as the basis for the respective specifications in Table 3.

In general, we find a positive and robust OARE effect that is statistically significant at the 1% level across all specifications. The marginal OARE effect ranges from +30% in column (3) to +44% in column (4). Notably, less productive institutions benefit substantially less from OARE registration than more productive institutions. This difference in OARE effects by productivity level is more pronounced when we examine the least productive and most productive institutions (see Section 5.3.2. below).

---

[27] Note that the Ranking Web of Universities that we use to create the rank variable is mainly based on the assessment of the web presence of institutions, e.g., it uses link analysis for quality evaluation. It is in this respect that an institution's web performance provides a proxy for its technical expertise to set up online access to journals.

[28] The effect of collaborations with authors from the US is positive and similar in size across specifications. It is statistically significant at the 5% level in column (8).

TABLE 3 | OARE EFFECT BY THE NUMBER OF QUARTERS WITH PUBLICATIONS

	(1)	(2)	(3)	(4)	(5)	(6)
# quarters with publication	<=25 quarters	<=25 quarters	<=25 quarters	>25 quarters	>25 quarters	>25 quarters
Model:	OLS with *distance_1*	OLS with *distance_2*	OLS with institution-discipline FE	OLS with *distance_1*	OLS with *distance_2*	OLS with institution-discipline FE
Dependent variable:	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)
OARE treated (DDD)	0.273***	0.273***	0.265***	0.362***	0.361***	0.351***
	(0.0389)	(0.0389)	(0.0382)	(0.0508)	(0.0508)	(0.0509)
# pages	-0.000304*	-0.000303*	0.00128	-7.86e-05	-7.79e-05	-1.59e-05
	(0.000158)	(0.000159)	(0.000831)	(0.00182)	(0.00182)	(0.00183)
# references	-7.52e-05	-7.46e-05	-0.000424	-0.000468	-0.000465	-0.000440
	(5.66e-05)	(5.67e-05)	(0.000263)	(0.000631)	(0.000632)	(0.000642)
# co-authors USA	-6.59e-05	-7.02e-05	-0.00314	0.00705	0.00694	0.00541
	(0.000590)	(0.000590)	(0.00277)	(0.00575)	(0.00574)	(0.00541)
# co-authors EUR	0.000616*	0.000615*	0.00390**	0.0135***	0.0135***	0.0131***
	(0.000374)	(0.000374)	(0.00179)	(0.00466)	(0.00466)	(0.00465)
# OARE references	0.000381***	0.000381***	-0.000170	-0.00116	-0.00116	-0.00124
	(0.000147)	(0.000147)	(0.000545)	(0.00144)	(0.00144)	(0.00146)
Rank2: 5,000<rank<=10,000	-0.0177	-0.0173		-0.104	-0.140	
	(0.0167)	(0.0167)		(0.360)	(0.364)	
Rank3: 10,000<rank<=15,000	-0.0199	-0.0196		-0.398*	-0.398	
	(0.0204)	(0.0204)		(0.239)	(0.253)	
Rank4: 15,000<rank<=25,000	-0.0188	-0.0185		0.0212	0.0342	
	(0.0157)	(0.0158)		(0.228)	(0.235)	
Rank5: rank=>25,000	-0.0398***	-0.0396***		-0.419***	-0.417***	
	(0.0115)	(0.0115)		(0.149)	(0.150)	
Pop1: 100<pop<=500, in 1,000	0.00808	0.00826		-0.399*	-0.446*	
	(0.00524)	(0.00524)		(0.238)	(0.237)	
Pop2: 500<pop<=1,000, in 1,000	0.0115*	0.0115*		-0.219	-0.249	
	(0.00624)	(0.00623)		(0.276)	(0.269)	
Pop3: 1,000<pop<=5,000, in 1,000	0.0131**	0.0128**		-0.223	-0.229	
	(0.00509)	(0.00531)		(0.221)	(0.207)	
Pop4: pop>5,000, in 1,000	0.00762	0.00776		-0.146	-0.137	
	(0.00551)	(0.00554)		(0.198)	(0.188)	
Distance from largest domestic city, in 100 km	-0.000346 (0.000378)			-0.0257 (0.0215)		
Distance from closest domestic city with > 1 million inh., in 100 km		-0.000405 (0.000437)			-0.0159 (0.0274)	
Constant	0.0711***	0.0696***	0.0377***	1.471***	1.373***	0.717***
	(0.0132)	(0.0130)	(0.00856)	(0.311)	(0.293)	(0.0305)
Quarter dummies	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	NO	YES	YES	NO
Institution-discipline FE	NO	NO	YES	NO	NO	YES
Observations	149,550	149,550	149,550	11,900	11,900	11,900
R-squared	0.0401	0.0401	0.0307	0.2181	0.2149	0.1259
Number of Inst_Discipline	2,991	2,991	2,991	238	238	238
Number of Inst	2,371	2,371	2,371	119	119	119

We use a balanced panel and take into account journal articles by both single and multiple local authors. Results on the impact of OARE membership (*treated*) on publication output of research institutions by the number of quarters with publications (<=25 quarters; >25 quarters) from OLS DDD estimation method. We use Stata's *xtreg* command. The institution-discipline-quarter pairs constitute the unit of observation. Period under study: 1^{st} quarter 2000 to 2^{nd} quarter 2012. We obtain the findings on the OARE effect mentioned in the text by calculating the exponential of the *treated* coefficient minus 1. Reference quarter is 36. Reference rank is *rank≤5000*. Reference population is *pop≤100*. Robust standard errors clustered at the institutional level.
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

TABLE 4 | OARE EFFECT (SINGLE LOCAL AUTHORS ONLY)

VARIABLES	(1) OLS Base	(2) + Article info	(3) +Rank	(4) +Popul.	(5) +Dist._1	(6) +Dist._2	(7) +Inst.-Disc. FE	(8) MCMC
Dependent variable:	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)
OARE treated (DDD)	0.339***	0.337***	0.332***	0.332***	0.332***	0.332***	0.322***	0.4402***
	(0.0322)	(0.0318)	(0.0315)	(0.0315)	(0.0315)	(0.0315)	(0.0311)	(0.0262)
# pages		0.000577	0.000596	0.000578	0.000577	0.000579	0.00164*	0.00007
		(0.000570)	(0.000574)	(0.000576)	(0.000576)	(0.000576)	(0.000880)	(0.0002)
# references		-0.000146	-0.000155	-0.000160	-0.000168	-0.000165	-0.000104	-0.0001*
		(0.000196)	(0.000196)	(0.000196)	(0.000196)	(0.000196)	(0.000274)	(0.00009)
# co-authors USA		0.00152	0.00171	0.00167	0.00161	0.00162	-0.000170	-0.0001
		(0.00311)	(0.00281)	(0.00282)	(0.00284)	(0.00283)	(0.00264)	(0.0009)
# co-authors EUR		0.00631	0.00612	0.00605	0.00608	0.00607	0.00825*	0.0036***
		(0.00385)	(0.00388)	(0.00389)	(0.00388)	(0.00388)	(0.00459)	(0.0008)
# OARE references		0.000754	0.000772*	0.000832*	0.000860*	0.000851*	0.000355	0.0006***
		(0.000461)	(0.000457)	(0.000458)	(0.000461)	(0.000462)	(0.000647)	(0.0001)
Rank2: 5,000<rank<=10,000			-0.235***	-0.228***	-0.226***	-0.225***		
			(0.0854)	(0.0849)	(0.0849)	(0.0853)		
Rank3: 10,000<rank<=15,000			-0.248***	-0.232***	-0.233***	-0.232***		
			(0.0791)	(0.0785)	(0.0782)	(0.0785)		
Rank4: 15,000<rank<=25,000			-0.0402	-0.0335	-0.0331	-0.0322		
			(0.104)	(0.104)	(0.103)	(0.103)		
Rank5: rank=>25,000			-0.320***	-0.325***	-0.324***	-0.324***		
			(0.0668)	(0.0671)	(0.0671)	(0.0671)		
Pop1: 100<pop<=500				0.0150	0.0172	0.0170		
				(0.0309)	(0.0314)	(0.0317)		
Pop2: 500<pop<=1,000				0.0579	0.0592	0.0584		
				(0.0365)	(0.0369)	(0.0368)		
Pop3: 1,000<pop<=5,000				0.0466	0.0388	0.0401		
				(0.0326)	(0.0312)	(0.0317)		
Pop4: pop>5,000				0.0494*	0.0404	0.0439		
				(0.0300)	(0.0292)	(0.0294)		
Distance from largest domestic city, in 100 km					-0.00225			
					(0.00166)			
Distance from closest domestic city with > 1 million inh., in 100						-0.00172		
						(0.00178)		
Constant	0.112***	0.105***	0.384***	0.338***	0.354***	0.343***	0.0502***	0.0215**
	(0.0165)	(0.0169)	(0.0660)	(0.0702)	(0.0696)	(0.0697)	(0.00946)	(0.0084)
ρ								-0.375***
								(0.0870)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO	NO
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES	YES
Observations	125,800	125,800	125,800	125,800	125,800	125,800	125,800	31,300
R-squared	0.085	0.082	0.129	0.131	0.131	0.131	0.066	-
Number of Inst_Discipline	2,516	2,516	2,516	2,516	2,516	2,516	2,516	626
Number of Inst	2,031	2,031	2,031	2,031	2,031	2,031	2,031	500

We use a balanced panel and take into account journal articles by single local authors only. Results on the impact of OARE membership (*treated*) on publication output of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru) from OLS DDD and Bayesian MCMC estimation methods. In (1) to (7), we use the *xtreg* command in STATA. In (8) we apply Bayesian MCMC estimation techniques. The institution-discipline-quarter pairs constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. We obtain the findings on the OARE effect mentioned in the text by calculating the exponential of the *treated* coefficient minus 1. Reference country is Nigeria. Reference quarter is 36. Reference rank is *rank≤5000*. Reference population is *pop≤100*. Robust standard errors clustered at the institutional level (OLS) and standard errors of the marginal posterior distributions (Bayesian) reported in parentheses. The MCMC results are shown for a random sample of 500 institutions. Bayesian estimation is based on the estimation of two equations and does not produce as such a value for R-squared.
$^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.3. Robustness

5.3.1. Articles with only one local author

In our basic model, we use a balanced panel including both single- and multiple-local-authored journal articles to examine the OARE effect. As discussed in Section 3.1, dividing authorship shares across different institutions is not a trivial exercise. For robustness, we therefore run the same regressions as reported in Table 2 for the subsample of single-local-author articles, i.e. articles for which we observe only one local author who may or may not be affiliated with an OARE member institution. As before, we use a balanced panel. We create the single-local-authored dataset by dropping 18,955 articles from the sample for which we have at least two local authors.

In Table 4, we use the same specifications as in Table 2. The marginal OARE effect is positive and statistically significant at the 1% level. It ranges from +38% in column (7) to +55% in column (8) where we use the Bayesian estimation method. These results provide additional empirical evidence for a robust OARE effect.

5.3.2. MCMC for different sub-samples

We use the Bayesian estimation method described in Appendix 1 to check how our results vary according to institutions that publish during different numbers of quarters. We divided the sample in 3 subsamples: institutions that published during less than 2 quarters, between 2 and 30 and during more than 30 quarters (out of 50 quarters).

The results reported in Table 5 show that the most productive institutions, i.e., institutions that published during more than 30 quarters, benefited the most from joining the OARE initiative (+57%). The marginal OARE effect is more than twice as large as the marginal OARE effect for institutions that published during 2 and 30 quarters (+26%), while institutions that published during less than 2 quarters barely benefited from joining the OARE initiative (+7%).

TABLE 5 | OARE EFFECT USING MCMC FOR DIFFERENT SUBSAMPLES

	(1)	(2)	(3)	(4)
Sample:	Full sample	*nquarter* < 2	2 ≤ *nquarter* ≤ 30	*nquarter* > 30
Dependent variable:	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)
OARE treated (DDD)	0.4505***	0.0683***	0.2326***	0.4517***
	(0.015)	(0.0186)	(0.0276)	(0.0283)
# pages	0.000004	0.0001	0.0029***	-0.001
	(0.0002)	(0.0002)	(0.0007)	(0.0019)
# references	-0.00009	-0.00001	-0.001***	-0.001*
	(0.00007)	(0.0001)	(0.0003)	(0.0006)
# co-authors USA	0.0019**	0.0092*	-0.001	0.0086*
	(0.0007)	(0.0047)	(0.0019)	(0.0049)
# co-authors EUR	0.0084***	0.0036	0.0035*	0.0102**
	(0.0008)	(0.0022)	(0.0020)	(0.0040)
# OARE references	0.0001	-0.0001	0.0008	0.0001
	(0.0001)	(0.0002)	(0.0005)	(0.0012)
ρ	-0.260***	0.0297	0.0150	-0.185***
	(0.0567)	(0.0373)	(0.0441)	(0.0465)
Quarter dummies	YES	YES	YES	YES
Institution-discipline FE	YES	YES	YES	YES
Number of observations	43,000	13,900	9,350	9,100
Number of Inst_Discipline	860	270	175	181
Number of Inst	650	250	125	125

We use a balanced panel and take into account journal articles by both single and multiple local authors. Sub-sample regressions are based on a random sample of institutions in columns (1), (2) and (3). Column (4) includes all institutions that publish during more than 30 quarters. Bayesian estimation is based on the estimation of two equations and does not produce as such a value for R-squared.
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

6. Conclusion

We have analyzed the effect of free and reduced-fee online access to the environmental science literature via the OARE initiative on scientific productivity in Bolivia, Ecuador, Kenya, Nigeria and Peru. We provide empirical support for a positive marginal OARE effect that ranges between +48% and +57%. The marginal OARE effect is also positive and statistically significant when we run the regressions separately for Group A (free access) and Group B (reduced-fee access) countries and for each of the five countries under study. In addition, a robustness check analyzing a balanced panel with single local authors yields qualitatively similar results, i.e. the marginal OARE effect ranges between +38% and +55%. Moreover, we use MCMC estimation examining different sub-groups to show the robustness of our results. Overall, our results provide empirical support for the hypothesis that free online access to journals promotes research in developing countries.

Nevertheless, we find that there is potential for improvement on two grounds. First, we found that the most productive institutions benefited the most from joining the OARE initiative while the least productive institutions barely benefited from joining. This result suggests that OARE increases the productivity difference between the most and least productive institutions. Under these conditions the least productive institutions are ceteris paribus less likely to catch up. Second, we find that not more than 5% of all eligible institutions joined OARE after a period of more than 5 years. This finding reveals the unused potential of the OARE initiative. Based on our results, policies aimed at increasing the awareness of free online access initiatives in developing countries should therefore be encouraged.

As a broader policy implication, our study suggests that an open access mandate or policy may promote scientific output – not only by research institutions in developing countries. Extending on the link between academic research and economic growth (see the literature discussed in section 2) our findings may hence point to direct economic effects as a higher research output level steaming from OARE membership may result in new environmental innovations. A natural follow up is to explore the question of whether OARE has increased the number of patent applications using free or reduced-fee access throughout the OARE program. In addition, it would be interesting to investigate in more detail how (open) online access has changed the way scientists do research and collaborate internationally.

Acknowledgements

The data used on this report were collected and main analyses done when Frank Mueller-Langer was Senior Research Fellow at Max Planck Institute for Innovation and Competition (MPI-IC). We are grateful to Kimberly Parker and the World Health Organization for providing us with data on registration dates of the institutions that joined the OARE initiative. We thank Dietmar Harhoff, Stan Liebowitz, Fabio Montobbio, Laura Rosendahl Huber, Michael Weber and the participants of the 2016 MPI-IC research seminar, the 2016 Workshop on "The Organization, Economics and Policy of Scientific Research", the 2015 MPI-IC guest lecture series and the 2015 SERCI Conference for comments. Research assistance was provided by Sebastian Osterrieth, Laura Sundsgaard and Christoph Winter.

References

Acs, Z. J., Audretsch, D. B., Feldman, M. P., 1994. R&D Spillovers and Recipient Firm Size. Review of Economics and Statistics, 76(2), 336-340.

Annan, K., 2004. Science for All Nations. Science, 303, 925.

Audretsch, D. B., Feldman, M. P., 1996. R&D Spillovers and the Geography of Innovation and Production. American Economic Review, 86(3), 630-640.

Bergstrom, T., Bergstrom, C. T., 2004. The Costs and Benefits of Library Site Licenses to Academic Journals. Proceedings of the National Academy of Sciences, 101(3), 897-902.

Bounide, D., Abel, F., Waelbroeck, P., 2016. Debit Card and Demand for Cash. Journal of Banking & Finance, 73, 55-66.

Dasgupta, P., David, P.A., 1994. Toward a New Economics of Science. Research Policy, 23(5), 487-521.

Dosi, G., 1988. Sources, Procedures, and Microeconomic Effects of Innovation. Journal of Economic Literature, 26(3), 1120-1171.

Eger, T., Scheufen, M., Meierrieks, D., 2015. The Determinants of Open Access Publishing: Survey Evidence from Germany. European Journal of Law and Economics, 39(3), 475-503.

European Commission, 2012. Towards Better Access to Scientific Information: Boosting the Benefits of Public Investments in Research. Brussels: European Commission.

Evans, J.A, Reimer, J., 2009a, Open Access and Global Participation in Science. Science, 323, 1025.

Evans, J.A., Reimer, J., 2009b, Open Access and Global Participation in Science, supporting online material. Science, 323.

Furman, J. L., Stern, S., 2011. Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. American Economic Review, 101(5), 1933-1963.

Gaulé, N., Maystre, N., 2011. Getting Cited: Does Open Access Help? Research Policy, 40(10), 1332-1338.

Geroski, P. A., 2000. Models of Technology Diffusion. Research Policy, 29(4-5), 603-625.

Griliches, Z., 1957. 'Hybrid Corn': An Exploration in the Economics of Technological Change. Econometrica, 25(4), 501-522.

Griliches, Z., 1992. The Search for R&D spillovers. Scandinavian Journal of Economics, 94, 29-47.

Hall, B., 2004. Innovation and Diffusion, NBER, Cambridge, MA NBER Working Paper No.10212.

Hall, B., Khan, B., 2003. Adoption of New Technology, NBER, Cambridge, MA NBER Working Paper No. 9730.

Hanauske, M., Bernius, S., Dugall, B., 2007. QuantumGame Theory and Open Access Publishing, Physica A., 382, 650-664.

Jaffe, A. B., 1989. Real Effects of Academic Research. American Economic Review, 79(5), 957-970.

Jeon, D.S., Rochet, J.C., 2010. The Pricing of Academic Journal: A Two-Sided Market Perspective. American Economic Journal: Microeconomics, 2, 222-255.

Mansfield, E., 1961. Technical Change and the Rate of Imitation. Econometrica, 29(4), 741-766.

Mansfield, E., 1963. Intrafirm Rates of Diffusion of an Innovation. Review of Economics and Statistics, 45(4), 348-359.

Merton, R.K., 1973. The Sociology of Science: Theoretical and Empirical Investigations, Storer, N.W. (editor), University of Chicago Press: Chicago, IL.

McCabe, M.J., Snyder, C.M., 2005. Open Access and Academic Journal Quality. American Economic Review Papers and Proceedings, 95(2), 453-458.

McCabe, M. J., Snyder, C. M, 2014. Identifying the Effect of Open Access on Citations Using a Panel of Science Journals. Economic Inquiry, 52(4), 1284-1300.

McCabe, M. J., Snyder, C. M., 2015. Does Online Availability Increase Citations? Theory and Evidence from a Panel of Economics and Business Journals. Review of Economics and Statistics, 97(1), 144-165.

Mueller-Langer, F., Scheufen, M., 2013. Academic Publishing and Open Access. In Towse, R. and Handke, C. (editors), Handbook of the Digital Creative Economy, pp. 365-377, Edward Elgar.

Mueller-Langer, F., Watt, R., 2018. How Many More Cites is a $3,000 Open Access Fee Buying You? Empirical Evidence from a Natural Experiment. Economic Inquiry, 56(2), 931-954.

Murray, F. E., Aghion, P., Dewatripont, M., Kolev, J., Stern, S., 2016. Of Mice and Academics: Examining the Effect of Openness on Innovation. American Economic Journal: Economic Policy, 8(1), 212-252.

Ramello, G. B., 2010. Copyright & Endogenous Market Structure: A Glimpse from the Journal-Publishing Market. Review of Economic Research on Copyright Issues, 7(1), 7-29.

Romer, P. M., 1986. Increasing Returns and Long-Run Growth. Journal of Political Economy, 94(5), 1002-1037.

Romer, P. M., 1990. Endogenous Technological Change. Journal of Political Economy, 98(5),

71-102.

Scheufen, M., 2015. Copyright versus Open Access: On the Organisation and International Political Economy of Access to Scientific Knowledge, Springer International Publishing.

Shavell, S., 2010. Should Copyright for Academic Works be Abolished? Journal of Legal Analysis, 2(1), 301-358.

Solow, R. M., 1956. A Contribution to the Theory of Economic Growth. Quarterly Journal of Economics, 70(1), 65-94.

Sorenson, O., Fleming, L., 2004. Science and the Diffusion of Knowledge. Research Policy, 33(10), 1615-1634.

Stephan, P.E., 1996. The Economics of Science. Journal of Economic Literature, 34(3), 1199-1235.

Stoneman, P., Battisti, G., 2010. The Diffusion of New Technology, Handbook of the Economics of Innovation, Volume 2, 733-760.

Suber, P., Arunachalam, S., 2005. Open Access to Science in the Developing World, http://dash.harvard.edu/bitstream/handle/1/4725025/suber_worldinfocity.htm?sequence=1

Suber, P., 2012. Open Access. Cambridge: MIT Press.

APPENDIX 1 | BAYESIAN METHODOLOGY

Equation (A1) determines the outcome of the endogenous binary variable:

$$y_{1,i} = \begin{cases} 1, if\ w_{1,i} > 0 \\ 0, if\ w_{1,i} \leq 0 \end{cases} \qquad \text{(A1)}$$

where $w_{1,i} = x_{1,i}'\ \beta_1 + \varepsilon_{1,i}$, β_1 is of dimension k_1 and $x_{1,i}$ is a set of k_1 control variables.

Equation (A2) explains the observed variable $w_{2,i}$ as a function of individual characteristics and the endogenous binary variable $z_{1,i} = 1$ if $w_{1,i} > 0$ and $z_{1,i} = 0$ if $w_{1,i} \leq 0$,

$$w_{2,i} = z_{1,i}\ \delta_1 + z_{2,i}'\ \delta_2 + \varepsilon_{2,i} = x_{2,i}'\ \beta_2 + \varepsilon_{2,i} \qquad \text{(A2)}$$

where δ_1 is the structural parameter associated with the binary endogenous variable z_1, $z_{2,i}$ is a set of k_2 explanatory variables not necessary identical to $x_{1,i}$ and δ_2 is a vector of parameters of dimension k_2, $x_{2,i} = (z_{1,i}, z_{2,i}')'$ and $\beta_2 = (\delta_1, \delta_2')'$.

We assume that $\varepsilon_i = (\varepsilon_{1,i}', \varepsilon_{2,i}')'$ is normally distributed with mean $(0, 0)'$ and covariance Σ for $i = 1, \ldots, n$: $\Sigma = \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}$. Parameter ρ represents the correlation between the unobservable variables. Parameter σ^2 is the variance of $\varepsilon_{2,i}$. Since the probit equation (A1) is not identified, we chose to normalize the variance of the endogenous binary variable to 1. This is a standard restriction in probit models.

Let $\beta = (\beta_1', \beta_2')'$, $w_1 = (w_{1,1}, \ldots, w_{1,n})'$, $w_2 = (w_{2,1}, \ldots, w_{2,n2})'$ and define $w = (w_1', w_2')'$. We define ε_1, ε_2, and ε in a similar fashion.

The covariance of the unobservable variables is simply

$$\Omega = E\varepsilon\varepsilon' = \Sigma \otimes I_n$$

where I_n denotes the identity matrix of dimension $n \times n$. Thus Ω^{-1} is readily obtained. We similarly define

$$X = \begin{bmatrix} x_{11} & 0 \\ 0 & x_2 \end{bmatrix} \quad 2n \times (k_1 + k_2)$$

The (partially) latent model can be written in matrix format:

$$w = X\beta + \varepsilon \qquad \text{(A3)}$$

Hence conditional on w and Ω, the estimates of β are simply obtained by a generalized least-squares (GLS) regression of (A3).[29] Moreover, the matrices $X'\Omega^{-1}X$ and $X'\Omega^{-1}w$ required for the GLS estimates of the parameters of the model are easily computed. We use a uniform prior for β, ρ and a non-informative prior for σ: $p(\beta, \rho, \sigma) \propto 1/\sigma$.[30] The Metropolis-Gibbs sampling algorithm proceeds in 4 steps drawing from conditional distributions sequentially. The full procedure is described in Bounie et al. (2016).

[29] Since each stage generally includes different sets of explanatory variables, we cannot estimate the seemingly unrelated regressions model with ordinary least-squares regression applied to each latent equation separately.

[30] The choice of the prior distribution does not matter much when there is a large number of observations. Moreover, using the uniform prior distribution provides a direct means of comparison with the maximum likelihood procedures.

APPENDIX 2 | SUMMARY STATISTICS BY COUNTRY GROUP

A. Summary statistics for Group A countries

VARIABLES	mean	sd	min	max	N
Dependent variable					
# publications	0.267	1.745	0	93.82	105,050
Countries					
Kenya	0.401	0.490	0	1	105,050
Nigeria	0.412	0.492	0	1	105,050
Bolivia	0.187	0.390	0	1	105,050
Main variable of interest					
OARE treated (DDD)	0.0110	0.104	0	1	105,050
Article characteristics					
# co-authors USA	0.579	1.288	0	36	105,050
# co-authors EUR	0.760	1.964	0	57	105,050
# OARE references	5.895	7.267	0	135	105,050
# pages	9.231	5.140	1	66	105,050
# references	30.09	17.99	0	293	105,050
Institutional characteristics					
Rank1: rank<=5,000	0.0171	0.130	0	1	105,050
Rank2: 5,000<rank<=10,000	0.0105	0.102	0	1	105,050
Rank3: 10,000<rank<=15,000	0.0228	0.149	0	1	105,050
Rank4: 15,000<rank<=25,000	0.0476	0.213	0	1	105,050
Rank5: rank=>25,000	0.902	0.297	0	1	105,050
rank, in 1,000	8.915	5.311	0.907	21.79	105,050
City characteristics					
Distance from largest domestic city, in 100 km	3.522	3.572	0	20.64	105,050
Distance from closest domestic city with more than 1 million inhabitants, in 100 km	1.979	2.866	0	20.64	105,050
Pop0: pop<=100, in 1,000	0.275	0.446	0	1	105,050
Pop1: 100<pop<=500, in 1,000	0.0866	0.281	0	1	105,050
Pop2: 500<pop<=1,000, in 1,000	0.220	0.414	0	1	105,050
Pop3: 1,000<pop<=5,000, in 1,000	0.382	0.486	0	1	105,050
Pop4: pop>5,000, in 1,000	0.265	0.441	0	1	105,050

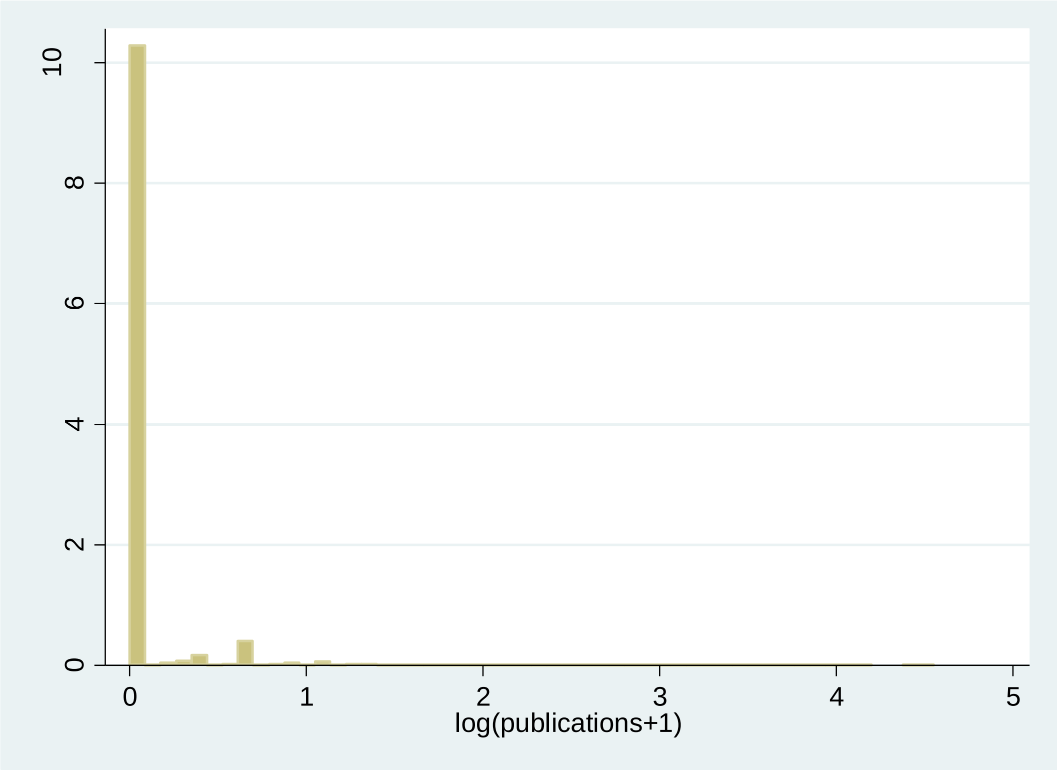
Data is aggregated at the institution-discipline-quarter level. The institution-discipline-quarter pairs constitute the unit of observation. We take into account journal articles by both single and multiple local authors in the Group A countries under study. Registered research institutions receive free OARE membership in Group A countries (GNI per capita below $1,600).

B. Summary statistics for Group B countries

VARIABLES	mean	sd	min	max	N
Dependent variable					
# publications	0.121	0.748	0	29.53	56,400
Countries					
Ecuador	0.351	0.477	0	1	56,400
Peru	0.649	0.477	0	1	56,400
Main variable of interest					
OARE treated (DDD)	0.00851	0.0919	0	1	56,400
Article characteristics					
# co-authors USA	1.006	2.013	0	37.25	56,400
# co-authors EUR	0.982	2.028	0	39	56,400
# OARE references	7.973	9.218	0	135	56,400
# pages	10.20	5.792	1	120	56,400
# references	34.89	19.06	0	213	56,400
Institutional characteristics					
Rank1: rank<=5,000	0.0488	0.215	0	1	56,400
Rank2: 5,000<rank<=10,000	0.0355	0.185	0	1	56,400
Rank3: 10,000<rank<=15,000	0.0186	0.135	0	1	56,400
Rank4: 15,000<rank<=25,000	0.00443	0.0664	0	1	56,400
Rank5: rank=>25,000	0.893	0.309	0	1	56,400
rank, in 1,000	4.036	1.770	0.749	21.39	56,400
City characteristics					
Distance from largest domestic city, in 100 km	2.754	3.668	0	15.73	56,400
Distance from closest domestic city with more than 1 million inhabitants, in 100 km	1.944	3.716	0	15.73	56,400
Pop0: pop<=100, in 1,000	0.145	0.352	0	1	56,400
Pop1: 100<pop<=500, in 1,000	0.148	0.355	0	1	56,400
Pop2: 500<pop<=1,000, in 1,000	0.0443	0.206	0	1	56,400
Pop3: 1,000<pop<=5,000, in 1,000	0.244	0.429	0	1	56,400
Pop4: pop>5,000, in 1,000	0.541	0.498	0	1	56,400

Data is aggregated at the institution-discipline-quarter level. The institution-discipline-quarter pairs constitute the unit of observation. We take into account journal articles by both single and multiple local authors in the Group B countries under study. Registered research institutions receive reduced-fee OARE membership ($1,000 per year) in Group B countries (GNI per capita below $5,000).

APPENDIX 3 | HISTOGRAM OF THE LOG NUMBER OF PUBLICATIONS



Histogram of log(number of publications +1) at the institution-discipline-quarter level. log(number of publications +1) is the dependent variable in the regressions.

APPENDIX 4 | OARE EFFECT (UNBALANCED PANEL)

Model:	(1) Base	(2) + Article info	(3) +Rank	(4) +Population	(5) +Dist._1	(6) +Dist._2	(7) Inst.-Disc. FE
Dependent variable:	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)	Log(*w+1*)
OARE treated (DDD)	0.756***	0.747***	0.518***	0.512***	0.513***	0.512***	0.300***
	(0.0954)	(0.0948)	(0.0650)	(0.0652)	(0.0650)	(0.0651)	(0.0390)
# pages		-0.000629	-0.000699	-0.000712	-0.000624	-0.000680	0.000652
		(0.00144)	(0.00140)	(0.00137)	(0.00138)	(0.00138)	(0.000915)
# references		-0.000359	-5.82e-05	-0.000112	-0.000232	-0.000150	-0.000261
		(0.000475)	(0.000462)	(0.000438)	(0.000438)	(0.000443)	(0.000309)
# co-authors USA		0.0140	0.0138*	0.0137*	0.0142*	0.0136*	-0.000213
		(0.0112)	(0.00817)	(0.00790)	(0.00806)	(0.00793)	(0.00335)
# co-authors EUR		0.0101	0.00695	0.00628	0.00667	0.00654	0.00748*
		(0.00831)	(0.00842)	(0.00844)	(0.00842)	(0.00838)	(0.00419)
# OARE references		0.00439***	0.00392***	0.00427***	0.00475***	0.00452***	-0.000349
		(0.00131)	(0.00136)	(0.00130)	(0.00132)	(0.00133)	(0.000702)
Rank2:			-0.241	-0.238	-0.222	-0.228	
5,000<rank<=10,000			(0.201)	(0.193)	(0.194)	(0.194)	
Rank3:			-0.344**	-0.299**	-0.307**	-0.297**	
10,000<rank<=15,000			(0.148)	(0.150)	(0.142)	(0.148)	
Rank4:			-0.0137	0.00485	0.00109	0.00698	
15,000<rank<=25,000			(0.167)	(0.160)	(0.157)	(0.159)	
Rank5: rank=>25,000			-0.395***	-0.422***	-0.419***	-0.419***	
			(0.0951)	(0.0954)	(0.0946)	(0.0951)	
Pop1: 100<pop<=500, in				-0.153	-0.146	-0.150	
1,000				(0.186)	(0.188)	(0.189)	
Pop2: 500<pop<=1,000,				0.0601	0.0491	0.0547	
in 1,000				(0.202)	(0.203)	(0.201)	
Pop3: 1,000<pop<=5,000,				0.00520	-0.0635	-0.0298	
in 1,000				(0.195)	(0.194)	(0.189)	
Pop4: pop>5,000, in				0.0177	-0.0490	-0.00875	
1,000				(0.188)	(0.192)	(0.185)	
Distance from largest					-0.0167**		
domestic city, in 100 km					(0.00853)		
Distance from closest dom. city						-0.00886	
with >1 million inh., in 100 km						(0.00943)	
Constant	0.840***	0.831***	1.128***	1.139***	1.258***	1.164***	1.246***
	(0.0596)	(0.0571)	(0.108)	(0.209)	(0.215)	(0.205)	(0.0245)
Quarter dummies	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	NO
Institution-discipline FE	NO	NO	NO	NO	NO	NO	YES
Observations	16,131	16,131	16,131	16,131	16,131	16,131	16,131
R-squared	0.1265	0.1343	0.1877	0.1956	0.1997	0.1962	0.0405
Number of Inst_Discipline	3,229	3,229	3,229	3,229	3,229	3,229	3,229
Number of Inst	2,490	2,490	2,490	2,490	2,490	2,490	2,490

We use an unbalanced panel and take into account journal articles by both single and multiple local authors. Results on the impact of OARE membership (*treated*) on publication output of research institutions in five developing countries (Bolivia, Ecuador, Kenya, Nigeria, Peru) from OLS DDD estimation methods. We use the *xtreg* command in STATA. The institution-discipline-quarter pairs constitute the unit of observation. Period under study: 1^{st} quarter 2000 to 2^{nd} quarter 2012. We obtain the findings on the OARE effect mentioned in the text by calculating the exponential of the *treated* coefficient minus 1. Reference country is Nigeria. Reference quarter is 36. Reference rank is *rank≤5000*. Reference population is *pop≤100*. Robust standard errors clustered at the institutional level.
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

APPENDIX 5 | OARE EFFECT BY COUNTRY GROUP

	(1)	(2)	(3)	(4)	(5)	(6)
Country Group:	Group A	Group A	Group A	Group B	Group B	Group B
Model:	OLS with *distance_1*	OLS with *distance_2*	OLS with inst.-disc. FE	OLS with *distance_1*	OLS with *distance_2*	OLS with inst.-disc. FE
Dependent variable:	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)
OARE treated (DDD)	0.447***	0.447***	0.440***	0.281***	0.281***	0.268***
	(0.0423)	(0.0423)	(0.0420)	(0.0511)	(0.0511)	(0.0489)
# pages	-0.000323	-0.000318	0.000208	0.000455	0.000454	0.00142
	(0.00104)	(0.00104)	(0.00128)	(0.000853)	(0.000853)	(0.00135)
# references	0.000184	0.000184	0.000303	-0.000383	-0.000383	-0.000628
	(0.000354)	(0.000355)	(0.000422)	(0.000302)	(0.000302)	(0.000469)
# co-authors USA	0.000629	0.000648	0.000935	0.00294	0.00294	0.000680
	(0.00279)	(0.00279)	(0.00327)	(0.00458)	(0.00458)	(0.00406)
# co-authors EUR	0.000780	0.000767	0.00133	0.0151***	0.0151***	0.0194***
	(0.00199)	(0.00199)	(0.00238)	(0.00582)	(0.00582)	(0.00604)
# OARE references	0.00130*	0.00129*	0.000982	-0.000116	-0.000108	-0.000227
	(0.000747)	(0.000747)	(0.000891)	(0.000723)	(0.000722)	(0.00106)
Rank2: 5,000<rank<=10,000	-0.144	-0.141		-0.311***	-0.310***	
	(0.220)	(0.220)		(0.0835)	(0.0835)	
Rank3: 10,000<rank<=15,000	-0.296*	-0.293*		-0.325***	-0.324***	
	(0.158)	(0.158)		(0.0774)	(0.0774)	
Rank4: 15,000<rank<=25,000	-0.0965	-0.0960		-0.334***	-0.332***	
	(0.168)	(0.168)		(0.0783)	(0.0785)	
Rank5: rank=>25,000	-0.418***	-0.417***		-0.337***	-0.336***	
	(0.148)	(0.148)		(0.0804)	(0.0804)	
Pop1: 100<pop<=500, in 1,000	0.0560	0.0555		0.00111	0.00185	
	(0.0397)	(0.0392)		(0.0140)	(0.0142)	
Pop2: 500<pop<=1,000, in 1,000	0.104**	0.101**		0.0221	0.0220	
	(0.0468)	(0.0471)		(0.0189)	(0.0189)	
Pop3: 1,000<pop<=5,000, in 1,000	0.0762**	0.0888**		0.0297	0.0281	
	(0.0367)	(0.0361)		(0.0204)	(0.0209)	
Pop4: pop>5,000, in 1,000	0.0731**	0.0818**		0.0409**	0.0390**	
	(0.0370)	(0.0362)		(0.0168)	(0.0168)	
Distance from largest domestic city,	-0.00255			-7.27e-05		
in 100 km	(0.00294)			(0.00131)		
Distance from closest domestic city with		0.00115			-0.000460	
more than 1 million inhabitants, in 100 km		(0.00397)			(0.00129)	
Constant	0.424***	0.403***	0.0812***	0.322***	omitted	0.0464***
	(0.150)	(0.148)	(0.0113)	(0.0738)		(0.0144)
Quarter dummies	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	NO	YES	YES	NO
Institution-discipline fixed effects	NO	NO	YES	NO	NO	YES
Observations	105,050	105,050	105,050	56,400	56,400	56,400
R-squared	0.1432	0.1430	0.0892	0.1481	0.1481	0.0496
Number of Inst_Discipline	2,101	2,101	2,101	1,128	1,128	1,128
Number of Inst	1,599	1,599	1,599	891	891	891

We use a balanced panel and take into account journal articles from both single and multiple local authors. Results on the impact of OARE membership (*treated*) on publication output of research institutions by country group (Group A: Bolivia, Nigeria, Kenya; Group B: Ecuador, Peru) from OLS DDD estimation. We use Stata's *xtreg* command. The institution-discipline-quarter pairs constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. We obtain the findings on the OARE effect mentioned in the text by calculating the exponential of the *treated* coefficient minus 1. Reference quarter is 36. Reference rank is *rank≤5000*. Reference population is *pop≤100*. Robust standard errors clustered at the institutional level.
**p* < 0.1, ***p* < 0.05, ****p* < 0.01.

APPENDIX 6 | OARE EFFECT BY COUNTRY

	(1)	(2)	(3)	(4)	(5)
Country:	Kenya	Nigeria	Bolivia	Ecuador	Peru
Model:	OLS +Inst.-Disc. FE	OLS +Inst.-Disc. FE	OLS +Inst.-Disc. FE	OLS +Inst.-Disc. FE	OLS +Inst.-Disc. FE
Dependent variable:	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)	Log(w+1)
OARE treated (DDD)	0.384***	0.466***	0.280***	0.209**	0.293***
	(0.0646)	(0.0532)	(0.0767)	(0.0877)	(0.0560)
# pages	0.000185	-0.000119	0.00118	0.00119	0.00173
	(0.00164)	(0.00246)	(0.00209)	(0.00220)	(0.00151)
# references	-0.000145	0.000564	0.000381	-0.00233***	7.37e-05
	(0.000506)	(0.000912)	(0.000598)	(0.000883)	(0.000501)
# co-authors USA	-0.00125	0.0120	0.00188	0.00100	-0.00404
	(0.00406)	(0.00862)	(0.00721)	(0.00353)	(0.00610)
# co-authors EUR	0.000753	0.00509	0.00184	0.0348***	0.0122**
	(0.00489)	(0.00923)	(0.00332)	(0.00832)	(0.00559)
# OARE references	0.00339***	0.00151	-0.00204*	0.00384*	-0.00177*
	(0.00122)	(0.00217)	(0.00108)	(0.00204)	(0.00107)
Constant	0.0671***	0.109***	0.0451**	0.0391	0.0493***
	(0.0163)	(0.0193)	(0.0218)	(0.0271)	(0.0172)
Quarter dummies	YES	YES	YES	YES	YES
Institution-Discipline FEs	YES	YES	YES	YES	YES
Observations	42,100	43,300	19,650	19,800	36,600
R-squared	0.0554	0.1106	0.0269	0.0562	0.0519
Number of Inst_Discipline	842	866	393	396	732
Number of Inst	637	641	321	324	567

We use a balanced panel and take into account journal articles from both single and multiple local authors. Results on the impact of OARE membership (*treated*) on publication output of research institutions by country from OLS DDD estimation. We use Stata's *xtreg* command. The institution-discipline-quarter pairs constitute the unit of observation. Period under study: 1st quarter 2000 to 2nd quarter 2012. We obtain the findings on the OARE effect mentioned in the text by calculating the exponential of the *treated* coefficient minus 1. Reference quarter is 36. Reference rank is *rank≤5000*. Reference population is *pop≤100*. Robust standard errors clustered at the institutional level.
*p < 0.1, **p < 0.05, ***p < 0.01.

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub