Journal of Scientometric Res. 2020; 9(2):154-162
http://www.jscires.org

Research Article

The USA Dominates World Research in Basic Medicine and Biotechnology

Ricardo Brito[1,*], Alonso Rodríguez-Navarro[1,2]

[1]Departamento de Estructura de la Materia, Física Térmica y Electrónica and GISC, Universidad Complutense de Madrid, Plaza de las Ciencias, Madrid, SPAIN.
[2]Departamento de Biotecnología-Biología Vegetal, Universidad Politécnica de Madrid, Avenida Puerta de Hierro, Madrid, SPAIN.

ABSTRACT

Objectives: Research is the foundation of technological progress; the aim of this study is to investigate research success in the USA, the EU and other countries in basic medical, biochemical and biotechnological topics. Methods: Research assessments were performed using the number of publications, the e_p index and the probability that a publication reaches the top 0.01% by citation. The e_p index reveals an important characteristic of citation distribution. The distribution of the publications from a country in global percentiles follows a power law and the e_p index is a mathematical derivative of the exponent of this power law. In addition to its intrinsic value as performance indicator, this index allows calculating a country's probability of publishing highly cited papers and, consequently, of achieving important discoveries or scientific breakthroughs. Findings: Our results show that the USA is scientifically ahead of all countries and that its research is likely to produce approximately 80% of the important global breakthroughs in the research topics investigated in this study. EU research has maintained a constant weak position with reference to USA research over the last 30 years. Countries different from the USA and the EU are increasing enormously their number of publications. Currently, the probability that these publications report an important breakthrough is slightly lower than that of the EU.

Keywords: Research evaluation, Percentile distribution, e_p index.

Correspondence

Ricardo Brito

Departamento de Estructura de la Materia, Física Térmica y Electrónica and GISC, Universidad Complutense de Madrid, Plaza de las Ciencias 3, 28040, Madrid, SPAIN.

Email: brito@ucm.es

Received: 12-05-2020
Revised: 15-07-2020
Accepted: 31-07-2020
DOI: 10.5530/jscires.9.2.19

INTRODUCTION

Research is the foundation of technological progress and countries and institutions require reliable research assessment methods to determine the profitability of their research investments. In the absence of reliable research assessments, the actual economic and societal benefits of research and its contribution to the progress of knowledge cannot be judged.

In a previous study we assessed the performance of technological research in three world geographical areas: the USA, the EU and the rest of the world.[1] However, this study only considered physical- and chemical-based technologies. Although these technologies support a large proportion of economic growth, two biological research fields, namely medicine and biotechnology, have a large societal and economic relevance. Currently, the Covid-19 epidemic makes manifest that research in these areas is of crucial importance; the present study is focused on these important research fields and aims to find out how the progress at the forefront of knowledge in these areas is globally distributed.

Medical research is very extensive and diverse;[2] our study is focused exclusively on basic medical research, which is strongly linked to biotechnology. In both of these fields knowledge progress is fast and a significant contribution to this progress requires research systems that are very competitive. Clinical and epidemiological studies are also of societal importance; however, they have not been included in this research because their study approaches[2] are different from those used in basic medical research and biotechnology.

Two methods have been used traditionally in research assessments: peer review and several types of metrics. Most rankings and cross-country research assessments are based on metrics and both types of assessments have been extensively applied to institutions; the results from these assessments allow comparative studies of the two methods. An extensive study based on the UK Research Exercise Framework (REF) concludes that metrics cannot be used for the substitution of peer review,[3] but two subsequent studies[4,5] have proved that this substitution is possible if some requirements are fulfilled. The most important of these requirements is that assessments must be performed at fairly high level of aggregation. In this case, different types of percentile-based metrics are reliable.

The aforementioned study of Traag and Waltman[4] uses the number of papers in the top 10% of global highly cited papers as the method of research assessment. This metric shows a high correlation with peer review in the REF2014,[6] but this type of dichotomous procedure "rely on the idea that only the upper part of the distribution matters"[7] providing a narrow picture of a complex distribution of citations. The inaccuracy comes from the fact that although the number of publications in the two upper parts of two particular citation distributions may be equal, the citations to these two sets of publications can be very different. To overcome this problem a non-dichotomous metric that reveals the functioning of the complete research system is necessary.

Non-dichotomous indicators

Non-dichotomous indicators subsume in a single parameter the research performance at all citation levels. Two non-dichotomous indicators with some similarities have been proposed and investigated: the I3[8,9] and the e_p[1] indices. The I3-index is calculated assigning weighting factors to percentile rank classes (e.g. top-1%, top-5%, top-10%, top-25%, top-50% and bottom-50%), which allows obtaining a single index that considers all published papers.

The e_p index is based on a mathematical property of the distribution of country's papers by percentiles in the citation-based global distribution of papers. The numbers of papers in top percentiles fit a power law, which implies that the exponent of this power law characterizes the functioning of the research system;[10] the e_p index is a simple derivative of the exponent of this power law (see Methods section). It was created in order that a higher e_p index indicates a higher performance because the exponent of the power law is lower when the performance is higher. In this study we used the e_p index because it is a non-dichotomous indicator and it does not require previous assumptions of percentile importance.

Methodology

As introduced above, our study is based on the percentile distributions of country's papers among global papers attending to citations counts. This distribution characterizes the competitiveness of country's research and can be measured by the e_p index.[10] The most competitive countries accumulate papers in small top percentiles (e.g. 0.01%, 0.1%, 1.0%) and exhibit a high e_p index and the opposite applies to the least competitive countries. The citation counts to perform the present study were taken from the Web of Science. The retrieved papers were ordered by their number of citations and analyzed as described below.

The procedure for the calculation of the e_p index has been described previously.[1,10,11] Briefly, global and country publications were ranked in parallel using the number of citations in decreasing order; the percentile limits in the global list (any percentile up to 100%) were fixed in according to the rank numbers and turned into the number of citations of the last paper of the selected percentile. Then, the country number of publications in each selected percentile was equal to the number of papers with the same or higher number of citations as the last paper of the selected percentile in the global list (i.e., the country ranking number of the last paper). When in the global and country lists the percentile limits occurred in sets of publications with the same number of citations,[12] the number of tied publications in the country set was fixed using a proportional method (i.e., proportional number of tied publications in global and country lists). This method is very accurate and can be used for those researches that cannot download hundreds of thousands of publications.

After counting the number of publications in a series of percentiles, we obtained the percentile distribution of the number of publications (i.e., cumulative frequencies) or the cumulative probability after dividing by the number of papers. The equations are:

$$N(x) = N \left(\frac{x}{100}\right)^{\alpha} \tag{1}$$

$$P(x) = \left(\frac{x}{100}\right)^{\alpha} \tag{2}$$

Where x is the selected top percentile and N is the total number of papers. The numbers of papers in percentiles were fitted to a power law as shown in Figure 1 (for statistical details see (11)). This figure shows the accuracy of this method in biological areas, which is similar to that found in technological areas[1] Figure 1. The e_p index is equal to 10 raised to minus the exponent of the fitted power law function.[11] The equations is:

$$e_p = 10^{-\alpha} \tag{3}$$

Where α is the exponent of the power law function. A country with an e_p index of 0.1 indicates that citations to the country's papers present the same distribution of those to world's papers. If the e_p index of a country is lower than 0.1, the research performance of that country or institution is worse than the global average. The cumulative probability calculated from e_p is:

$$P(x) = e_p^{(2 - \lg x)} \tag{4}$$

Where x is the selected percentile. In the present study, the top 0.01 percentile was selected for the reasons given elsewhere.[1] The expected frequency of papers in the top 0.01% of the most cited publications is equal to the cumulative probability, $P(x)$, multiplied by the number of publications. This cumulative frequency is named $P_{top\ 0.01\%}$ using the Leiden Ranking notation.[13,14]

It is worth noting that the three parameters are highly related but different. The e_p index is both percentile- and size-independent and reveals the *breakthrough potential* of the research system.[1] The other two parameters depend on the selected percentile; the probability is size independent and the cumulative frequency is size dependent.

Topics

The purpose of this study was to investigate the country's research performances across the world in basic biology and its application to biotechnology and medicine. To perform it we used the Web of Science Core Collection Advanced Search feature; to select the terms of the search query we took into consideration that country's research performance varies enormously not only across research fields (e.g., physics, chemistry, or biology), which seems logical, but also depending on the research activity of the field, (i.e., hot or quiescent topics.[10,15-17] Therefore, we focused in research areas and topics that are currently highly investigated.

For basic biology and biotechnology, the selection of the query terms was simple and could be restricted to two Research Areas of the Web of Science Core Collection available for the Advanced Search (SU=): Biochemistry and Molecular Biology and Biotechnology and Applied Microbiology. The adequacy of these research areas to our purpose was clear for the scientific coincidence and because they include very highly cited research topics. In fact, among the 22 Research Fields of the Essential Science Indicators, the field Molecular Biology and Genetics, which is in the basis of the selected Research Areas, has the highest number of citations per paper of the 22 Research Fields.

The construction of the search query for basic medical research required more consideration because none of the Research Areas that can be included in the Advanced Search query fulfilled our requierements. Therefore, we constructed the query with hot topics (TS=) in basic medical research; to select these topics we attended to several conditions: they had to be (i) basic; (ii) highly cited; and (iii) extensive, in order that the number of retrieve papers were high. To find these topics, firstly we retrieved the papers published in *Nature*, *Science* and the *Proceedings of the National Academy of Sciences USA* and ordered them by their number of citations. Then we selected the most frequent biological topics that fulfilled our conditions and checked, one by one, that they were highly cited among global publications including all journals. From the results we selected: cancer, microbiota, stem cell*, immunity and inflamma* (the asteric denotes truncation). Although these topics do not include all possible highly cited topics in basic medical research they are the most important and a representative sample of this type of research. In fact, their number of publications was high, 179,951 in 2018, almost the same number than in the whole research area of Chemistry, 190,290. The annual number of publications on individual topics was different; cancer and inflamma* represente almost 60% and 30%, respectively, of the total number of publications. The topic CRISPR was identified only in recent years; highly cited papers were published starting in 2007 and they were frequent since 2013. Although the effect of this topic was restricted to the last year of our study, we included it for its current high biological importance and certain increas of significance in future studies.

Countries

This study has two parts: the evolution of the two selected research fields in the period 1984-2014 and the state of the two fields across countries in 2014. For the first part, we divided the world (we consider the 50 most productive countries in 2014) into three geographical research areas: the ERA (European Research Area), the USA and Others (i.e., all countries excluding the ERA countries and the USA). These areas were analyzed independently, omitting collaborative publications between them. For the second part, the eight largest and scientifically most active countries in the ERA: Germany, France, UK, Italy, Spain, Sweden, The Netherlands and Switzerland were studied. In Others, we selected the six most active countries in the selected research topics and areas: Australia, Canada, China, Japan, South Korea and Taiwan. In some cases instead of the ERA countries we record results for the EU excluding the UK in order to cancel out the dominant role that Switzerland and the UK play in ERA research.[18]

In the 15 selected countries we retrieved only domestic papers (all authors in the same country or set of countries; see discussion below). Some collaborative publications between two countries were also studied to complement the study of domestic papers.

Bibliometric searches

Bibliometric searches were performed in the Science Citation Index Expanded of the Web of Science Core Collection (WoS), using the "Advanced Search" feature. In each search we included a year (PY=), a country or set of countries (CU=) and a set of topics (TS=) or research areas (SU=), as described below.

For highly cited basic medical research we used: TS=((cancer OR crispr* OR micro biota OR stem cell* OR immunity OR inflamma*) NOT (statistics OR trial OR survey)) and for biochemistry and biotechnology we used: SU=((biochemistry and molecular biology OR biotechnology and applied biotechnology OR cell biology OR microbiology) NOT (computer science OR mathematical and computational biology)) NOT TS=(cancer OR crispr* OR micro biota OR stem cell* OR immunity OR inflamma* OR statistics OR trial

OR survey). We retrieved only "Articles," which excludes review papers, because in many cases review papers receive more citations than the original articles in which they are based.[19] Searches were performed between February 23 and March 5, 2018. Some countries were analyzed in different days but each analysis in a different day was complete including world and country citation distributions.

RESULTS

Basic medical research

To obtain a first overview of world research in basic medical topics, we studied the evolution of the three aforementioned parameters between 1984 and 2014 (Figure 2). The number of publications in the selected medical topics increased enormously over these 30 years, from 3,686 to 148,375 annual publications. Growth in the USA and the ERA was similar, while in others it was much higher, especially over the last 10 years (2004–2014). Throughout this time, the research performance in the USA, as revealed by the e_p index and the paper probability of reaching the top 0.01 percentile, was much higher than in the ERA and Others. The performance of the USA was already the best in 1984 and it increased over time. The e_p index showed a clear tendency of the ERA catching up with the USA in the 1984–1994 period; however, this tendency decreased in the 1994–2004 period and disappeared in 2004–2014. Even in 1984–1994, the paper probability of reaching the top 0.01 percentile showed that the ERA was clearly lagging compared to the USA. This apparently contradictory trend between the e_p index and paper probability of reaching the top 0.01 percentile was due to the original great difference between the USA and the ERA and the mathematical relationship between both parameters, which is exponential. Because of the original large difference, the ERA should increase its e_p index more than the USA to decrease the difference in paper probability of reaching the top 0.01 percentile; for example, if the USA ($e_p \approx 0.10$) increased the e_p index by 10% the ERA ($e_p \approx 0.06$) should increase it by 25% to similarly increase the probability.

Due to the large increase in the number of publications, the cumulative frequency of publications in the top 0.01 percentile ($P_{top\ 0.01\%}$) increased in the three geographical areas studied here; however, because of the higher US research performance, the difference between the USA and both the ERA and others increased permanently over the 30-year period.

Next, we calculated the research performance parameters for individual countries (Table 1). The countries with the highest e_p index values were Switzerland (0.15) and the USA (0.12), both of which were above the world average. All the other countries performed worse than the global average (< 0.10); the e_p index varied from approximately 0.08 in the UK, Netherlands, Germany and Sweden to less than 0.05 in South Korea, Japan and Taiwan. The performance of the EU excluding the UK was similar to that of Canada and Australia. Among the four biggest EU countries (Germany, Spain, France and Italy) Germany and France performed better than the EU excluding the UK. The probability of publishing a paper in the 0.01 percentile reflected the differences in the e_p index and the expected frequency of papers in the top 0.01% of the most cited publications, $P_{top\ 0.01\%}$, reflected both the differences in the e_p index and in the number of publications. The $P_{top\ 0.01\%}$ indicator was 10-fold higher in the USA than in the EU excluding the UK and 5-fold higher than in the ERA.

To further investigate the dominant role of USA research, we calculated the performance parameters of collaborations between the USA and others countries (we actually studied co-authorship.[20] The results summarized in Table 2 reveal that any country collaborating with the USA substantially improved its e_p index and paper probability of reaching the top 0.01 percentile. However, the increase in this probability was irregular; for example, it amounted 38-fold in Germany but only 2.7-fold in Switzerland. The increase of the probability was so high that although the number of collaborative papers was much lower than the number of domestic papers, the collaborative $P_{top\ 0.01\%}$ indicator was much higher than for domestic papers—with the exception of Switzerland. With reference to domestic papers, the collaboration between Switzerland and Germany increased 11-fold the probability for a paper from Germany to reach the 0.01 percentile but there was no increase for Switzerland's papers.

Apparently, the USA also benefited from some of these collaborations, especially with Germany, Canada and Switzerland, as they increased the domestic USA e_p index; however, this is not the best explanation as we discuss below.

Biochemistry and biotechnology

In this case, the evolution of the number of publications over 30 years was substantially different from that observed in basic medical research (compare Figures 2 and 3). The number of USA and ERA publications increased in the 1984–1994 period and remained constant or decreased after this time. Only others retained a permanent growth throughout the 30-year period. It is worth noting that these searches were performed using research areas that were made up by a collection of journals. Therefore, the differential growth in the number of papers between areas is highly dependent on the inclusion of new journals in them and changes in topic preferences by researchers.

Despite the differences in the evolution of the number of publications in basic medical research and biochemistry and biotechnology, the evolution of the e_p index in these two research fields showed the similarity of small changes, positive evolution in the USA and neutral or negative evolution in the other two geographical areas. The other size independent indicator, the probability that a paper reach the top 0.01 percentile, evolved as the e_p index because it is mathematically linked to it. As observed in basic medical research, also in biochemistry and biotechnology, the performance of the USA was already the best in 1984 and it increased over time.

The analysis of countries (Table 3) showed that the USA, the UK, Switzerland and the Netherlands performed clearly better than the world average (e_p index > 0.1). The performance by Sweden and Australia was approximately equal to the world average (e_p index = 0.1) and all the other countries showed worse performances (e_p index < 0.1). The EU excluding the UK showed rather a poor performance. As with basic medical research, South Korea, Taiwan and Japan showed the lowest performances, which were considerably below the average global performance. The probability of publishing a paper in the 0.01 percentile reflected the differences in the e_p index and the $P_{top\ 0.01\%}$ reflected both the differences in the e_p index and in the number of publications. The $P_{top\ 0.01\%}$ indicator was 12-fold higher in the USA than in the EU excluding the UK and 4-fold higher than in the ERA.

Once again, collaboration with the USA substantially improved the research performance of all countries tested (Table 4); the increase in paper probability of reaching the top 0.01 percentile was very irregular. For example, it was more than 300-fold higher for Japan. For Switzerland and Germany the collaboration with the USA increased the top 0.01%

Table 1: Research performance parameters in hot basic medical topics, in 15 selected cases. Publications in 2014; domestic counts.

Countries	Number of publications	e_p index	Paper probability for top 0.01%	$P_{(top\ 0.01\%)}$
Switzerland	734	0.1479	0.0004782	0.3510
USA	28818	0.1180	0.0001936	5.5778
UK	3398	0.0822	0.0000456	0.1550
Netherlands	1721	0.0814	0.0000439	0.0755
Germany	4817	0.0806	0.0000423	0.2038
Sweden	1094	0.0791	0.0000392	0.0429
France	3175	0.0721	0.0000270	0.0856
Canada	2705	0.0670	0.0000202	0.0546
EU w/o UK	27241	0.0665	0.0000196	0.5331
Australia	2155	0.0644	0.0000172	0.0370
Spain	2419	0.0579	0.0000113	0.0272
China	23602	0.0566	0.0000103	0.2426
Italy	4231	0.0543	0.0000087	0.0369
South Korea	5751	0.0484	0.0000055	0.0315
Japan	7925	0.0470	0.0000049	0.0387
Taiwan	2682	0.0331	0.0000012	0.0032

Table 2: Research performance parameters in hot basic medical topics, in seven cases of research collaborations between two countries. Publications in 2014.

Countries	Number of publications	e_p index	Paper probability for top 0.01%	$P_{top\ 0.01\%}$
Germany and USA	767	0.2001	0.001604	1.2303
Canada and USA	1059	0.1919	0.001356	1.4365
Switzerland and USA	206	0.1904	0.001315	0.2708
Japan and USA	696	0.1442	0.000433	0.3013
China and USA	3481	0.0990	0.000096	0.3343
South Korea and USA	774	0.0846	0.000051	0.0397
Switzerland and Germany	261	0.0812	0.000044	0.0114

Table 3: Research performance parameters in Biochemistry and Biotechnology research areas, in 15 selected cases. Publications in 2014; domestic counts.

Country	Number of publications	e_p index	Paper probability for top 0.01%	$P_{(top\ 0.01\%)}$
USA	12542	0.1323	0.0003068	3.847
UK	1795	0.1317	0.0003009	0.540
Switzerland	499	0.1219	0.0002210	0.110
Netherlands	652	0.1208	0.0002132	0.139
Sweden	470	0.1011	0.0001043	0.049
Australia	936	0.0938	0.0000775	0.073
Germany	2782	0.0827	0.0000468	0.130
Canada	1557	0.0771	0.0000354	0.055
China	10991	0.0740	0.0000301	0.330
France	1793	0.0677	0.0000211	0.038
EU w/o UK	15212	0.0662	0.0000192	0.292
Italy	1563	0.0534	0.0000081	0.013
Spain	1638	0.0505	0.0000065	0.011
South Korea	2379	0.0489	0.0000057	0.014
Taiwan	895	0.0418	0.0000031	0.003
Japan	3597	0.0375	0.0000020	0.007

Table 4: Research performance parameters in biochemistry and biotechnology research areas, in seven research collaborations between two countries. Publications in 2014.

Countries	Number of publications	e_p index	Paper probability for top 0.01%	$P_{top\ 0.01\%}$
Switzerland and USA	112	0.2600	0.0045709	0.512
Japan and USA	285	0.1654	0.0007482	0.213
Germany and USA	409	0.1654	0.0007482	0.213
Canada and USA	403	0.1438	0.0004278	0.172
Switzerland and Germany	127	0.1303	0.0002887	0.037
China and USA	1403	0.1157	0.0001791	0.251
South Korea and USA	329	0.1062	0.0001273	0.042

probability by 21- and 16-fold, respectively. The collaboration between Switzerland-Germany increased 6-fold the top 0.01% probability of Germany but insignificantly the probability of Switzerland. Although the number of publications resulting from these collaborations was lower than the number of domestic papers, the cumulative frequency of publications in the top 0.01 percentile was higher in collaborations. Only in Germany did the improved research performance not compensate for the lower number of publications.

Apparently, the USA also benefited from some of these collaborations, especially with Switzerland, as they increased the domestic USA e_p index; however, as aforementioned, this is not the best explanation as we discuss below.

DISCUSSION

To quantify research performance in countries and institutions, we have used three parameters derived from citation distributions: the e_p index, the probability for a country publication to reach the global top 0.01 percentile and the cumulative frequency of publications in this percentile. As explained previously, these parameters are mathematically calculated from the country's distribution of publications in the global publication percentiles based on citations. The e_p index is percentile-independent while the other two parameters require a specific percentile to be selected.[1]

The rationale behind the method of selecting a low percentile to estimate research performance is the assumption that the number of highly cited publications correlates with the number of discoveries or breakthroughs that a research system produces (Ref. [10] and references therein). This number of important discoveries or breakthroughs that boost science

Table 5: $P_{top\ 0.01\%}$ per million inhabitants across countries. Data obtained from the results presented in Tables 1 and 3.

Country	Basic medical topics	Biochem and Biotechnol
Switzerland	0.04387	0.01379
USA	0.01743	0.01202
Netherlands	0.00444	0.00818
Sweden	0.00429	0.00490
Germany	0.00249	0.00159
UK	0.00235	0.00818
Australia	0.00154	0.00302
Canada	0.00148	0.00149
France	0.00132	0.00058
Spain	0.00065	0.00025
Italy	0.00062	0.00022
South Korea	0.00062	0.00027
Japan	0.00030	0.00006
China	0.00017	0.00024
Taiwan	0.00013	0.00011

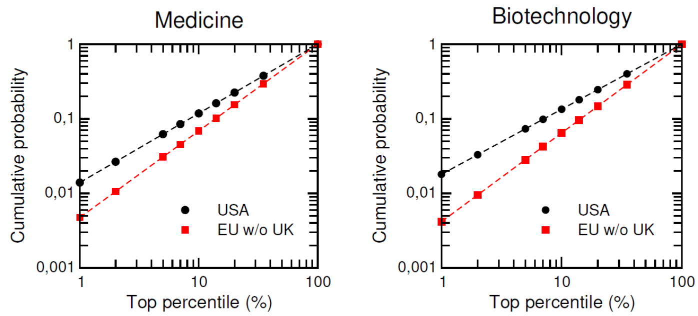

Figure 1: Cumulative probability plots for a paper from the USA and the EU excluding the UK to reach a specific top percentile. Publications in 2014.

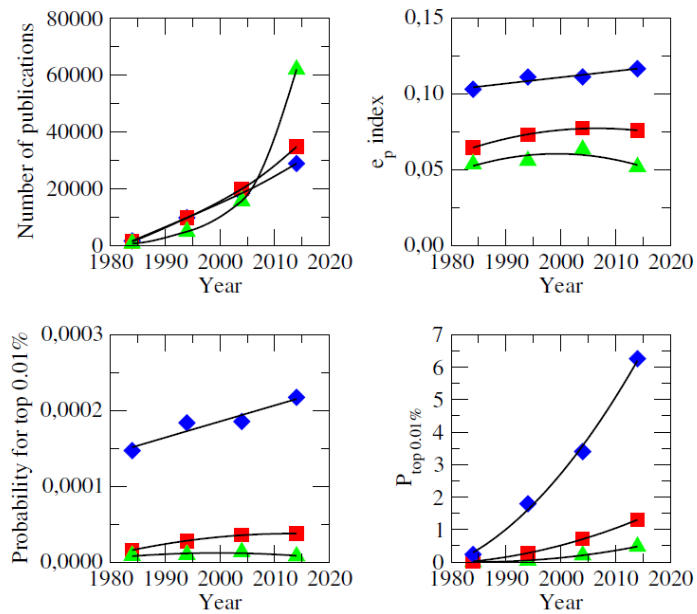

Figure 2: Evolution of research performance in the USA, the European Research Area (ERA) and other countries in basic medical research over 30 years. Curves are drawn to guide the eye. Symbols: diamonds, the USA; squares, the ERA; triangles, other countries.

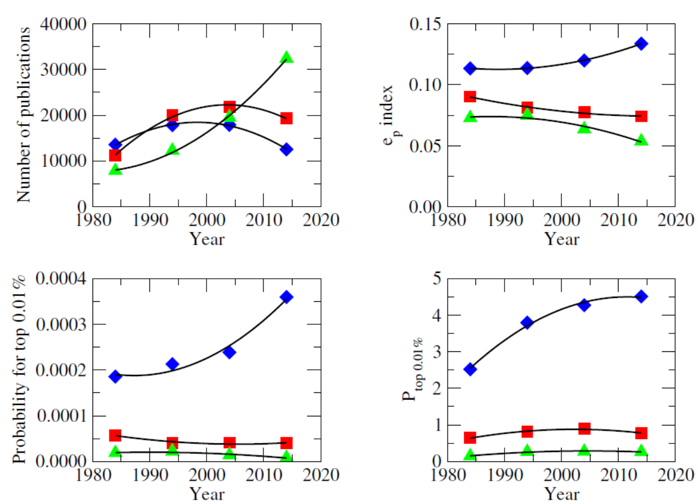

Figure 3: Evolution of research performance in the USA, the European Research Area (ERA) and other countries in biochemistry and biotechnology over 30 years. Curves are drawn to guide the eye. Symbols: diamonds, the USA; squares, the ERA; triangles, other countries.

and breakthrough innovations is very low, as is the number of the highly cited papers that report these achievements. In consequence research performance has to be evaluated at a low percentile. The low number of breakthrough papers precludes counting these papers; however, the probability of its achievement can be calculated.[11] It is worth highlighting that the correlation between important achievements and the high number of citations does not imply that one of the parameters measures the other and that, therefore, it cannot be applied to low aggregation levels such as individual researchers or small groups.[21-23] In conclusion, the probability that a paper published in a country reaches a low percentile reveals the country's probability of achieving important breakthroughs or discoveries of a similar frequency.

The selection of the top 0.01 percentile implies that in the topics and research areas studied, which produce approximately 150,000 and 80,000 annual publications, respectively, a total of 15 and 8 important discoveries or breakthroughs would be expected per year, which seems reasonable.[24] Considering less important discoveries or breakthroughs a higher top percentile would be used, e.g., 0.1 or even 1.0. This less stringent percentile selection would reduce the country differences revealed by the probability and expected frequency at the top 0.01 percentile.

The basic medical topics we study here are quite different from the fast evolving technological topics we studied previously,[1] which belong to physical and chemical fields. However, in both cases, citation levels are similarly high, which reveals high interest of researchers for these research fields. This similarity may have been the reason why country research assessments in the two quite different scientific fields (biology versus physics and chemistry) show many coincidences; however, for reasons we have not investigated they also show notable differences. The coincidence is the large advantage of the USA over the ERA and the difference is that the others are approaching to the USA in terms of the expected number of very highly cited publications in fast evolving technological but not in basic medical research. Regarding the number of important discoveries and breakthroughs, our results suggest that the USA produces approximately 40% of them in fast evolving technological topics (Table 3)[1] and approximately 80% in the hot basic medical topics (Figure 2). However, it is important to note that this remarkable high share of breakthroughs was obtained when publishing a quarter of the global number of papers in 2014.

The biochemical and biotechnological research areas studied here are not as highly cited as the basic medical topics; however, they are still highly cited. In these areas the USA also shows a clear leadership publishing approximately 80% of the very highly cited papers but only 20% of the total number of papers (Figure 3).

Further studies are necessary to determine the relationship between R&D investments, the total number of publications and the indicators used in this study. In any case, the large differences across countries in research performance attending to the probability of a paper reaching the top 0.01 percentile (Tables 1 and 3) strongly suggest that knowledge production depends more on research performance than on the total number of publications or R&D investments and capital. Therefore, the use of the last two parameters in econometric studies is necessarily misleading, because it is equivalent to considering a constant research performance across countries. We agree with the notion that "the benefits of scientific discovery have been heavy-tailed" [25] but this and previous[1] studies show that the heavy-tailed discoveries vary enormously across countries. Therefore, it can be concluded that the benefits of research are likely to be highly variable across countries and that they can be high or low independently of R&D investments or capital, at least in economically advanced countries.

Here we have studied the research performance of the ERA, the EU and other countries counting only domestic papers (i.e., excluding collaborations with external countries). This method does not measure the total scientific production of the research actors; however, in our opinion, it is the only method that can reveal the actual research performance level of a country or group of countries. Many arguments have been given in support of different counting methods,[26] but no method has been developed or can be developed to distribute unbalanced knowledge contributions among collaborating countries in unbalanced international collaborations.[27] For example, the data in Tables 1–4 strongly suggest that considering collaborations between USA and China would mistakenly improve the research performance of Chinese researchers. It is worth noting that this apparent improvement of research performance that obtains a partner with a low research performance does not occur if the partner is a highly competitive country such as Singapore in fast evolving technological topics.[1]

Another issue is the apparent improvement in research performance that also occurs for the USA in collaborative papers; collaboration with Germany, Canada, or Switzerland increased the USA e_p index by 1.6-fold in basic medical topics (Tables 1 and 2) and 2.0-fold in the collaboration with Switzerland in biochemistry and biotechnology (Tables 3 and 4). However, this increase might not imply a real increase of the *breakthrough potential* of the USA. The most probable reason for this increase is that these international collaborations do not occur at random with all USA research institutions but occurs preferentially with top institutions in which the e_p index is much higher than the USA e_p index, which is a national average. For example, the e_p index of the

Massachusetts Institute of Technology (MIT) or Harvard University in the two research fields studied here is almost 0.3 (unpublished results), which is similar to that measured in the aforementioned collaboration between USA and Switzerland. It is hard to believe that collaborations between MIT or Harvard University and Switzerland or Germany would improve the e_p index of MIT and Harvard University.

International collaborations (actually co-authorships)[20] are numerous[28,29] and raise a complex problem in research evaluation and policy. This problem does not fall under the scope of this study and the conclusions above apply exclusively to the research topics and the research fields studied here.

On average, the probability that a paper reaches the top 0.01 percentile is much lower when the paper is published in EU continental countries than when it is published in the USA (approximately 10- and 16-fold in hot technological topics and biochemistry and biotechnology, respectively). Aside from the discussion of whether this countries' ratio accurately reflects the ratio for the achievement of discoveries and breakthroughs, the empirical fact cannot be denied and raises the question of why EU research is scarcely successful in publishing highly cited papers. The e_p index in the most successful countries in continental Europe, Switzerland and the Netherlands, is similar to that of the USA; however, there are differences that raise some doubts about the comparison. Switzerland and the Netherlands are small countries with a low number of research universities that maintain a high level of research performance. In contrast, due to its size, the USA has many universities. In top USA universities the e_p index is around 0.3 (unpublished results); however, at the end of the ranking the figures might be 100-fold lower.

From the point of view of the economy of a country, it seems logical that for a similar effect bigger countries will need higher numbers of scientific breakthroughs. To relate country size and research success the cumulative frequency for the top 0.01 percentile ($P_{top0.01\%}$) can be divided by the GDP or the number of inhabitants. Using the latter normalization in basic medical topics (Table 5), the first country is Switzerland, very prominent and the second country is the USA; Netherlands and Sweden are the next two countries. In biochemistry and biotechnology, again the first two countries are Switzerland and the USA. The UK and Netherlands are the next two countries

Irrespective of the measurement method, it is evident that the four biggest continental EU countries keep a low research performance if the USA is taken as a reference, which raises a question about the causes. A key clue to answering this question might be that the greater differences occurs in fast evolving research topics[1,15-17] that arouse greater interest in society and researchers. It is remarkable that in the WoS research areas "plant sciences" and "physiology" the USA and the EU are similarly successful. The e_p index has not been calculated in these areas, but the similarity of the double rank plots (Figure 1 and 2)[30] allows the prediction of very similar e_p index values. This situation is puzzling, suggesting differences in researchers' motivations.

CONCLUSION

Our results based on the number (size-dependent indicator) and probability (size-independent indicator) of very highly cited papers (top 0.01%) and on the e_p index show the global leadership of USA research in basic medicine and biochemistry and biotechnology. In 2014, although the USA published roughly 20% of the global number of papers, it published 80% of the top 0.01% papers. Considering size, Switzerland shows a higher ratio than the USA in top 0.01% papers per inhabitant, but all other countries show lower ratios.

In our 30-year study of the evolution of the scientific performance of the USA and the EU suggests that scientific competence and efficiency is an intrinsic characteristic of old research systems that does not change easily. In this period, the strong USA research has evolved slightly stronger while the weak EU research undergoes minimal positive or negative fluctuations.

ACKNOWLEDGEMENT

This work was supported by the Spanish Ministerio de Economía y Competitividad (Grant number FIS2017-83709-R).

CONFLICT OF INTEREST

The author(s) declare no conflict of interest.

REFERENCES

1. Rodríguez-Navarro A, Brito R. Technological research in the EU is less efficient than the world average: EU research policy risks Europeans' future. Journal of Informetrics. 2018;12(3):718-31.
2. Röhrig B, DuPrel JB, Wachtlin D, Blettner M. Types of study in medical research. Part 3 of a series on evaluation of scientific publications. Deutsches Ärzteblatt International. 2009;106(15):262-8.
3. Wilsdon J, Allen L, Belfiore E, Campbell P, Curry S, Hill S, *et al.* The metric tide: Report of the independent review of the role of metrics in research assessment and management. Sage. 2016.
4. Traag VA, Waltman L. Systematic analysis of agreement between metrics and peer review in the UK REF. Palgrave Communications. 2019;5(1):29.
5. Rodríguez-Navarro A, Brito R. Like-for-like bibliometric substitutes for peer review: advantages and limits of indicators calculated from the ep index. Research Evaluation. 2020;29(2):215-30.
6. REF2014. Assessment framework and guidance on submissions. REF 02.2011. Hefce, Northavon House Coldharbour Lane, Bristol. 2011.
7. Albarrán P, Herrero C, Ruiz-Castillo J, Villar A. The Herrero-Villar approach to citation impact. Journal of Informetrics. 2017;11(2):625-40.
8. Leydesdorff L, Bornmann L. Integrated impact indicators compared with impact factors: An alternative research design with policy implications. Journal of the American Society for Information Science. 2011;62(11):2133-46.
9. Leydesdorff L, Bornmann L, Adams J. The integrated impact indicator revised (I3*): A non-parametric alternative to the journal impact factor. Scientometrics. 2019;119(3):1669-94.
10. Brito R, Rodríguez-Navarro A. Research assessment by percentile-based double

rank analysis. Journal of Informetrics. 2018;12(1):315-29.

11. Rodríguez-Navarro A, Brito R. Probability and expected frequency of breakthroughs: Basis and use of a robust method of research assessment. Scientometrics. 2019;119(1):213-35.
12. Waltman L, Schreiber M. On the calculation of percentile-based bibliometric indicators. Journal of the American Society for information Science and Technology. 2013;64(2):372-9.
13. Bornmann L, Wagner C, Leydesdorff L. BRICS countries and scientific excellence: A bibliometric analysis of most frequently cited papers. Journal of the Association for Information Science and Technology. 2015;66(7):1507-13.
14. Waltman L, Calero-Medina C, Kosten J, Noyons ECM, Tijssen RJW, Eck NJV, *et al*. The Leiden ranking 2011/2012: Data collection, indicators and interpretation. Journal of the American Society for information Science and Technology. 2012;63(12):2419-32.
15. Bonaccorsi A. Explaining poor performance of European science: institutions versus policies. Science and Public Policy. 2007;34(5):303-16.
16. Rodríguez-Navarro A, Narin F. European paradox or delusion-Are European science and economy outdated?. Science and Public Policy. 2018;45(1):14-23.
17. Sachwald F. Europe's twin deficits: Excellence and innovation in new sectors. Luxembourg: Publications Office of the European Union. 2015.
18. Rodríguez-Navarro A, Brito R. Might Europe one day again be a global scientific powerhouse? Analysis of ERC publications suggests it will not be possiblewithout changes in research policy. Quantitative Science Studies. 2020;1(2):872-93.
19. Miranda R, Garcia-Carpintero E. Overcitation and overrepresentation of review papers in the most cited papers. Journal of Informetrics. 2018;12(4):1015-30.
20. Katz JS, Martin BR. What is reserach collaboration?. Research Policy. 1997;26(1):1-18.
21. Allen L, Jones C, Dolby K, Walport M. Looking for landmarks: the role of expert review and bibliometric analysis in evaluating scientific publication outputs. Plos One. 2009;4(6):e5910.
22. Ruiz-Castillo J. The evaluation of citation distribution. SERIEs. 2012;3:291-310.
23. Raan AFJV. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics. 2005;62(1):133-43.
24. Bornmann L, Ye A, Ye F. Identifying landmark publications in the long run using field-normalized citation data. Journal of Documentation. 2018;74:278-88.
25. Press WH. What's so special about science (and how much should we spend on it?). Science. 2013;342(6160):817-22.
26. Gauffriau M. A categorization of arguments for counting methods for publication and citation indicators. Journal of Informetrics. 2017;11(3):672-84.
27. Zanotto SR, Haeffner C, Guimaraes JA. Unbalanced international collaboration affects adversely the usefulness of countries' scientific output as well as their technological and social impact. Scientometrics. 2016;109(3):1789-814.
28. Leydesdorff L, Wagner CS, Park HW, Adams J. International collaboration in science: The global map and the network. El Profesional de la Información. 2013;22:87-94.
29. Wagner CS, Park HO, Leydesdorff L. The counting growth of global cooperation netwoks in research: A conumdrum for national gevernments. Plos One. 2015;10(7):e0131816.
30. Rodríguez-Navarro A, Brito R. Double rank analysis for research assessment. Journal of Informetrics. 2018;12(1):31-41.