

# The Citation Merit of Scientific Publications

Juan A. Crespo<sup>1\*</sup>, Ignacio Ortuño-Ortín<sup>2</sup>, Javier Ruiz-Castillo<sup>2</sup>

<sup>1</sup> Departamento de Economía Cuantitativa, Universidad Autónoma de Madrid, Madrid, Spain, <sup>2</sup> Departamento de Economía, Universidad Carlos III de Madrid, Madrid, Spain

## Abstract

We propose a new method to assess the merit of any set of scientific papers in a given field based on the citations they receive. Given a field and a citation impact indicator, such as the mean citation or the  $h$ -index, the merit of a given set of  $n$  articles is identified with the probability that a randomly drawn set of  $n$  articles from a given pool of articles in that field has a lower citation impact according to the indicator in question. The method allows for comparisons between sets of articles of different sizes and fields. Using a dataset acquired from Thomson Scientific that contains the articles published in the periodical literature in the period 1998–2007, we show that the novel approach yields rankings of research units different from those obtained by a direct application of the mean citation or the  $h$ -index.

**Citation:** Crespo JA, Ortuño-Ortín I, Ruiz-Castillo J (2012) The Citation Merit of Scientific Publications. PLoS ONE 7(11): e49156. doi:10.1371/journal.pone.0049156

**Editor:** Santo Fortunato, Aalto University, Finland

**Received:** May 24, 2012; **Accepted:** October 9, 2012; **Published:** November 13, 2012

**Copyright:** © 2012 Crespo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors acknowledge financial support from the Santander Universities Global Division of Banco Santander. JR-C also acknowledges financial help from the Spanish MEC (Ministerio de Educación) through grant SEJ2007-67436. JAC and IO-O also acknowledge financial help from the Spanish MEC through grant ECO2010-19596. This paper is produced as part of the project Science, Innovation, Firms and markets in a Globalised World (SCIFI-GLOW), a Collaborative Project funded by the European Commission's Seventh Research Framework Programme, Contract number SSH7-CT-2008-217436. Any opinions expressed here are those of the author(s) and not those of the European Commission. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: juan.crespo@uam.es

## Introduction

The scientific performance of a research unit (a university department, research institute, laboratory, region, or country) is often identified with its publications and the citations they receive. There are a variety of citations-based specific indices for assessing the impact of a set of articles. Among the most prominent are the mean citation and the  $h$ -index, but there are many other possibilities. Regardless of the citation impact indicator used, the difficulty of comparing units that produce a different number of papers—even within a well-defined homogenous field—must be recognized. To better visualize the problem consider a concrete example. Suppose that we use as indicator the mean citation. Consider the articles published in Mathematics in 1998 and the citations they receive until 2007. The mean citation of papers published in Germany and Slovenia are 5.5 and 6.4, respectively. However, Germany produced 1,718 articles and Slovenia only 62. According to the mean citation criterion the set of Slovenian articles has greater relative impact than the German set. We will see, however, that according to the novel proposal introduced in this paper the performance exhibited by Germany has greater merit than that of Slovenia. No doubt this is an extreme example, but highlights a general difficulty that is present when comparing research units producing a different number of papers in the same field. Furthermore, we show that this difficulty in comparing sets of different sizes persists even if they are large. Thus, the problem in our example is not due to the small number of papers published in Slovenia. This difficulty is even more apparent for citation impact indicators that are size dependent, such as the  $h$ -index [1,2].

Comparisons across fields are even more problematic. Because of large differences in publication and citation practices, the

numbers of citations received by articles in any two fields are not directly comparable. Of course, this is the problem originally addressed by relative indicators recommended by many authors [3–11]. A convenient relative impact indicator is the ratio between the unit's observed mean citation and the mean citation for the field as a whole. Thus, after normalization, mean citations of research units in heterogeneous fields become comparable [12]. However, we argue that, as in the previous example of Germany and Slovenia, comparisons using normalized mean citations do not capture the citation merit of different set of articles.

The main aim of this paper is to propose a method to measure the citation merit of a set of articles a research unit publishes in a homogeneous field over a certain period. It should be clarified at the outset that the merit is conditional on the indicator used (mean,  $h$ -index, median, percentage of highly cited papers, etc.) and on the set of articles used as reference, which we will call “population of interest” (usually all the world articles published in a field in a given period). Thus, a given set of papers in a certain field and time period may have different merit depending on the citation impact indicator used. Given a citation impact indicator, our method allows for comparisons between sets of papers of different sizes and fields. Thus, we will be able to make statements like “The scientific publications of Department X in field A have a greater citation merit than the publications of Department Y in field B”.

Our method is based on a very simple and intuitive idea. Given a field and a citation impact indicator, the merit of a given set of  $n$  articles is identified with the percentile in which its observed citation impact lies on the distribution of citation impact values corresponding to all possible subsets of  $n$  articles in that field. Suppose, for example, that the impact indicator is the mean citation, and that the population of interest is equal to

all articles published in the world in a certain period in that field. In this case, the merit of a given set of  $n$  papers is given by the probability that a randomly drawn set of  $n$  articles in that field has a lower mean citation. Note that, since the merit of a set of papers of a research unit is associated with a percentile (or a probability), it is possible to compare two such percentiles for research units of different sizes working in different fields.

This method resembles that used in other areas such as, for example, Pediatrics where the growth status of a child is given by the percentile in which his/her weight lies within the weight distribution for children of the same age. In our case “same age” is equivalent to “same number of articles”. There is, however, an essential difference: in our case we do not compare the performance of a given research unit with  $n$  articles with the performance of other existing research units with a similar number of articles, but with the distribution generated by all possible subsets of  $n$  articles from a given pool of articles.

A related idea that also distinguishes between citation impact and citation merit can be found in [13] for the evaluation of scientific excellence in geographical regions or cities. The citation impact indicator they use is the percentage of articles in a city that belong to the top-10% most-highly cited papers in the world. As they say “the number of highly-cited papers for a city should be assessed statistically given the number of publications in total”. Thus, the scientific excellence of a city depends on the comparison between its observed and its expected number of highly cited papers.

The  $h$ -index has become very popular because it can be seen as capturing both quantity and quality. The original proposal by Hirsch [14] was designed for the evaluation of individual researchers, but it can be easily extended to research units (A research unit has  $h$ -index  $h$  if  $h$  of its articles have at least  $h$  citations each, and the remaining articles have no more than  $h$  citations each). However, due to its dependence on the number of articles, research units that have more articles also tend to have higher  $h$ -index values. For the different institutions they study, Molinari and Molinari [1] show that an universal relation emerges across institutions that enable them to empirically decompose the  $h$ -index as a product of a size independent factor  $h_m$  and a size dependent one  $N^{0.4}$ , where  $N$  is the number of papers. This factor  $h_m$  is then used among others by Kinney [2] who compares the scientific output of several U.S. institutions. In our case, we do not need to rely on any empirical estimation of the relation between  $h$ -index and the size  $N$  (moreover our methodology can be applied to all citation impact indicators).

In order to implement our method, a large dataset with information about world citation distributions in different homogeneous fields is required. In most of this paper, we use a dataset acquired from Thomson Scientific, consisting of all articles published in 1998–2007, and the citations they received during this period. We show that our approach yields rankings of research units quite different from those obtained by a direct application of the mean citation and the  $h$ -index.

## Methods

Consider a homogeneous scientific field (for example, Nuclear Physics, Molecular Biology, etc.). Suppose that we want to compare the relative merit of two sets of articles  $X$  and  $Y$ . Denote by  $x = \{x_1, \dots, x_i\}$  the vector of citations received over a fixed period by the  $i$  articles in  $X$ , and by  $y = \{y_1, \dots, y_j\}$  the corresponding vector of citations for the  $j$  articles in  $Y$ . Denote by

$W$  the set of articles used as a “population of interest”, and by  $w = \{w_1, \dots, w_N\}$  the vector of citations of the  $N$  articles in  $W$ . We require that  $X, Y \in W$ . In most applications in the paper we take  $W$  as the set of all articles published in the world in a given year in that field.

We next need some citation impact indicator  $g(\cdot)$  such as, for example, the mean citation or the  $h$ -index. The mean citation is perhaps the most often-used indicator, but recently the  $h$ -index has also become popular. Our method is silent about which is the most appropriate citation impact indicator. Given an indicator, we could compare  $x$  and  $y$ 's impact by comparing the numbers  $g(x)$  and  $g(y)$ . As indicated in the Introduction, such a direct comparison has important drawbacks and is often misleading. Thus, we propose a way to compare the merit of any two vectors of citations using the information  $g(x)$ ,  $g(y)$ ,  $i$ ,  $j$ , and  $w$ .

Denote by  $C_n$  the set of all subsets of  $W$  of size  $n$ . We take  $C_n$  as our sample space and the corresponding  $\sigma$ -algebra is given by all the subsets of  $C_n$ , i.e.  $\mathcal{A} = 2^{C_n}$ . We establish a probability function  $P: \mathcal{A} \rightarrow \mathbb{R}_+$  satisfying that all the simple events are equiprobable, i.e.

$$P(x_1) = P(x_2) = \frac{1}{\|C_n\|} \text{ for all } x_1, x_2 \in C_n.$$

where  $\|C_n\|$  denotes the (finite) number of elements in the set  $C_n$ , i.e.  $\|C_n\| = \binom{N}{n}$ .

Given the measure space  $(C_n, \mathcal{A}, P)$  we define the random variable  $g_n: C_n \rightarrow \mathbb{R}$  which is just our chosen impact indicator  $g$  restricted to sets of  $n$  articles. The cumulative distribution function (CDF) of  $g_n$ ,  $G_n: \mathbb{R} \rightarrow [0, 1]$ , is defined by.

$$G_n(z) = P(g_n^{-1}(-\infty, z]).$$

Note that  $G_n(z)$  denotes the probability that a subset of  $n$  articles from  $W$  has a vector of citations  $r = \{r_1, \dots, r_n\}$  such that  $g(r) \leq z$ .

**Definition.** The citation merit of a set of  $n$  papers  $X$  with citation vector  $x = \{x_1, \dots, x_n\}$  is given by  $G_n(g(x))$ . We write  $q_n(x) = G_n(g(x))$ .

Thus, we associate the citation merit of  $x = \{x_1, \dots, x_n\}$  with the percentile in which the number  $g(x)$  lies in the distribution  $G_n$ .

It should be emphasized that to determine the merit of a set of articles we just have to calculate a percentile (or probability) which does not require any statistical inference exercise.

In many cases we know the analytical expression of the function  $G_n(z)$ . For instance, in the case of the  $h$ -index, the function  $G_n(z)$  can be calculated exactly as described by Molinari and Molinari ([1], Equations A3, A6). This is a combinatorial formula that only requires to know the vector of citations in the population of interest,  $w = \{w_1, \dots, w_N\}$ .

However, in other instances, it might be difficult to calculate the analytical expression of  $G_n(z)$ . In these cases, one could approximate  $G_n(z)$  by taking  $s$  random draws of  $C_n$ . As in many empirical applications the cardinality of  $C_n$  is large, the number of draws should be large (in our applications we use 100,000). Thus whenever it is difficult to compute the combinatorial formula that gives the exact value of  $G_n(z)$  we proceed as follows: Let  $x_i = \{x_{i1}, \dots, x_{in}\}$ ,  $i = 1, \dots, s$ , be the vector of citations obtained in the  $i$ -th draw. Apply the impact indicator to each of these  $s$  sets and denote by  $g_n = \{g(x_1), \dots, g(x_s)\}$  the resulting vector. Let  $\hat{G}_n$  be the distribution function associated to such vector, so that  $\hat{G}_n(z)$  gives the percentage of components in vector  $g_n$  with a value equal

or less than  $z$ . Given a database with the information of  $w$ , this is a feasible and simple approach to approximate the probability  $q_n(x)$ .

To further motivate our citation merit definition, think of the following hypothetical example. Consider a given field and period and suppose that each article has only one author, and each author has written only one article. Suppose that the research unit is a university department that has published  $n$  papers, obtaining a citation impact level equal to  $g(x)$ . Suppose that instead of the actual department composition the chair could hire  $n$  persons from the pool of world researchers who have written a paper in the same field, and let  $x'$  be the corresponding vector of citations. Assume that the chair of the department hires these  $n$  people in a random way (so there is no difference from what a monkey would do). What would the probability be that  $g(x')$ , the citation impact level associated with such hypothetical random hiring, is lower than the actual value  $g(x)$ ? Such probability is our citation merit value  $G_n(z)$ .

Coming back to the example presented in the Introduction about the articles in Mathematics of Slovenia and Germany and judging by their mean citation of 6.3 and 5.5, Slovenia ranks higher than Germany. However, the merit values we obtain for the sets of papers of these two countries are 85.30 and 97.00, respectively. The probability that a set of 62 papers, randomly chosen from the pool of all papers published in Mathematics, have a mean lower than 6.3 is 85.30%, whereas the probability that a set of 1,718 papers have a mean lower than 5.5 is 97.00%. Thus, although the mean citation for Slovenia is higher than the mean citation for Germany, its merit is lower.

It is important to note that the result in this example is not just due to the fact that the “sample size” for Slovenia is very small (62 papers). Our empirical results provide many similar type of examples for larger number of papers. Consider, for example, the field of Engineering. Taiwan has 1,882 articles and mean citation of 5.58. Scotland has 610 articles and basically the same mean citation, 5.54. However, the merit of these two sets of articles are 31.20 and 36.00, respectively. Thus, even in cases with “large” number of articles our merit function might rank sets of articles differently from the rank obtained by the mean citation.

Given a citation impact indicator and a population of interest, the method just introduced allows us to compare sets of articles in the same field, and rank all of them in a unique way. Moreover, since the merit definition is associated with a percentile in a certain distribution, we can also make meaningful merit comparisons of sets of articles from different fields.

## Results

We use a dataset acquired from Thomson Scientific, consisting of all publications in the periodical literature appearing in 1998–2007, and the citations they received during this period. Since we wish to address a homogeneous population, in this paper only research articles are studied. After disregarding review articles, notes, and articles with missing information about Web of Science category or scientific field, we are left with 8,470,666 articles. For each article, the dataset contains information about the number of citations received from the year of publication until 2007 (see [15], for a more detailed description of this database).

We only consider two citation impact indicators: the mean citation, and the  $h$ -index. As already indicated, in the case of the  $h$ -index, our merit function  $G_n(z)$  can be calculated analytically. In the case of the mean, the precise combinatorial formula for  $G_n(z)$  is complicated and, in practice, it is not feasible given the large size of our datasets. Then we use the approximation  $\hat{G}_n$  described above.

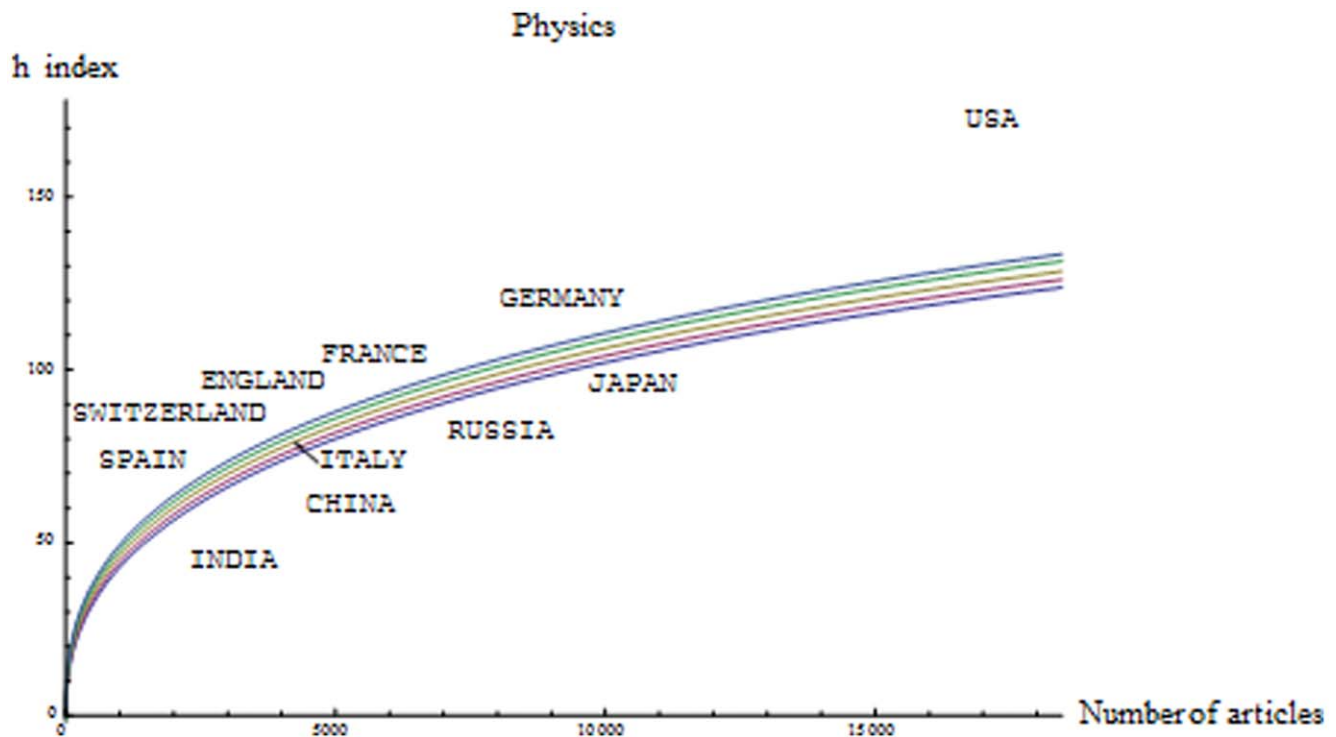
Since the mean and the standard deviation of  $W$  are known, one could think of approximating  $G_n(z)$  using the Central Limit Theorem, at least for research units with large numbers of articles. However, for all scientific fields the distribution of  $w$  is heavily skewed [15–19], and the underlying distribution might not have a finite variance, so that the Central Limit Theorem could fail even for research units with a large number of articles. (We have indeed checked that for the scientific fields used in the paper, and the sample sizes given the number of papers published by the research units considered, the distribution of the means of random samples is far from a normal distribution).

## Countries

In a first exercise, research units are countries, and the homogeneous fields are identified with the 22 broad fields distinguished by Thomson Scientific, 20 in the natural sciences, and two in the social sciences. In an international context we must confront the problem raised by cooperation between countries: what should be done with articles written by authors belonging to two or more countries? Although this old issue admits different solutions (see [20] for a discussion), in this paper we side with many other authors in following a multiplicative strategy (see [21–23]). Thus, in every internationally co-authored article a whole count is credited to each contributing area. Excluding the Multidisciplinary field, for each of the remaining 21 fields we compute the citation merit of the papers of each country according to the mean citation and the  $h$ -index, taking as the population of interest all papers published in the world in the corresponding field. We exclude the Multidisciplinary field because of the high heterogeneity of some of its journals. In doing so we exclude many high-impact articles published in, for example, Nature, Science and PNAS. One should incorporate such articles to their corresponding fields. This is, however, a laborious task that is beyond the scope of this paper. Figure 1 illustrates an example of our methodology when citation impact is measured by the  $h$ -index for the articles published in 1998 in the field of Physics, their citations until 2007, and a selection of countries. For each different value of  $n$ , Figure 1 shows the value of the  $h$ -index corresponding to percentiles 10, 25, 50, 75 and 90 of the corresponding distribution  $G_n$ , as well as the number of articles published by each country and its associated  $h$ -index.

Note that by just observing the  $h$ -index of, for example, Japan, France, Germany, and Switzerland, it is difficult to assess their relative merit. The reason, of course, is that the  $h$ -index is highly dependent on the number of articles. Thus, since Japan (9,600 articles), France (6,056), and Germany (9,598) produce more articles than Switzerland (2,028), they also have a higher  $h$ -index. However, with our method we are able to compare the publications of these countries using  $q_n(x)$ , the percentile where the observed  $h$ -index lies. It turns out that obtaining by chance an  $h$ -index as high as the one of Switzerland -with 2,028 papers- is a much more “unlikely” event than obtaining the  $h$ -index of any of the other three countries with their corresponding number of articles. Thus, our method assigns more merit to Switzerland (basically percentile 100) than to Japan (percentile 5.40), France (percentile 99.00), and Germany (percentile 99.90). Figure 1 also shows that the U.S. produces the largest number of articles, has the highest  $h$ -index and, according to our methodology, basically reaches the 100 percentile. This is a feature that appears in most of the 21 fields that we have analyzed.

Table 1 continues with the case of articles published in Physics in 1998 and equivalent tables for the remaining 20 fields are available as supplementary information to this paper, Tables S1. For the forty countries with the largest production, the tables



**Figure 1. Field of Physics.** Papers published in 1998. Curves represent for each number of articles the corresponding value of the  $h$ -index for which the merit is 10.00, 25.00, 50.00, 75.00 and 90.00 respectively. Points represent the number of papers and the values of the  $h$ -index for different countries.

doi:10.1371/journal.pone.0049156.g001

provide the  $h$ -index, the mean citation, the corresponding  $q_n(x)$  values, and the confidence intervals for our approximation in the case of the mean citation. For example, Italy has an  $h$ -index of 81, the sixth highest value in our sample. But if we look at the merit index  $q_n(x)$ , it falls to the eleventh position. It is observed that any of the two impact indices and its corresponding merit index  $q_n$  produce different rankings. There are many examples where the discrepancy between the two is very large. Thus, our methodology delivers outcomes that are quite different from those obtained by the direct use of the mean citation or the  $h$ -index criterion.

In some cases our methodology cannot discriminate enough between countries with very high merit indices. Consider for example the case of Clinical Medicine in Table 2, where Column 4 shows the merit index for a selection of countries when the citation impact is measured by the  $h$ -index. All these countries, except Germany, have a very similar merit index close to 100%. The reason for this result is that we are using as a population of interest all articles published in the world, and the quality of the articles published by this selection of countries is much higher than that of the rest of the world. Therefore, there are not many corresponding subsets of articles with citation impact as high as those observed in the countries in question. One possible way to discriminate among these “very high quality” sets of articles is to take as population of interest,  $W^*$ , only articles published in these countries. Column 5 in Table 2 shows the citation merit index in this case. Notice that when  $W$  contains all the papers published in the world France reaches the 99.4% percentile. However, in the case of  $W^*$ —a set of papers of a much higher quality than the  $W$  set—basically about half of all subsets of size 13,822 have an  $h$ -index higher than the one of France (140). Thus, in this case France’s percentile is 55.3%. Notice that changing the population of interest might produce a re-ranking of the citation merit. When

$W$  is used, England obtains a higher citation merit than Belgium. However, the opposite is the case when the population of interest is  $W^*$ . This possibility of re-ranking is not surprising since our notion of merit is based on the comparison of the observed  $h$ -index with the probability of obtaining sets of articles with lower  $h$ -indices. Such probability depends on the distribution function associated to the population of interest.

### University Departments and Laboratories

It could be argued that the broad fields so far analyzed are, in effect, too heterogeneous, a fact that may well diminish the value of our results. In this subsection we present comparisons of the merit of the publications of some selected university departments and laboratories in two more homogeneous scientific sub-fields (Thomson Scientific assigns articles to 219 Web of Science categories through the journals where they have been published, but many journals are simultaneously assigned to two or more categories. For a discussion of the strategies to deal with the problem raised by the multiple assignments of articles to Web of Science categories, see [24]). Tables 3 and 4 show the performance of some institutions in the sub-fields of Neuroscience and Economics, respectively (the data on the papers published by members of these departments has been obtained from the Web of Science of Thomson Scientific). The tables show the number of papers, the  $h$ -index, the mean citation, and the corresponding  $q_n(x)$ .

As before, there are significant discrepancies between the ranking according to the direct citation impact indicator ( $h$ -index or mean citation) and our merit function  $q_n(x)$ . Notice that many departments get a value of  $q_n(x)$  equal or very close to 100%. As already explained in the case of Clinical Medicine in Table 2, this is not surprising since all of them are top departments and the

**Table 1.** Field of Physics.

Country	Articles ( <i>n</i> )	<i>h</i> -index		Mean citation	
		<i>h</i> -index	Percentile ( <i>q<sub>n</sub></i> )	Mean citation	Percentile ( <i>q<sub>n</sub></i> )
USA	18267	173	100.00	19.52	100.00 (100.00, 100.00)
JAPAN	9600	99	5.40	11.96	36.71 (36.411, 37.008)
GERMANY	9598	114	99.90	15.54	100.00 (100.00, 100.00)
RUSSIA	8116	75	0.00	7.49	0.00 (0.000, 0.000)
FRANCE	6056	97	99.00	15.10	100.00 (100.00, 100.00)
ENGLAND	4890	90	99.30	14.68	100.00 (100.00, 100.00)
CHINA	4294	59	0.00	7.80	0.00 (0.000, 0.000)
ITALY	4086	81	87.70	14.40	99.99 (99.987, 99.998)
INDIA	2239	45	0.00	8.40	0.00 (0.000, 0.000)
SPAIN	2089	67	98.40	15.15	99.94 (99.929, 99.958)
SWITZERLAND	2028	81	100.00	21.96	100.00 (100.00, 100.00)
SOUTH KOREA	1911	51	0.00	9.77	0.00 (0.000, 0.000)
CANADA	1895	61	76.40	14.98	99.86 (99.841, 99.887)
POLAND	1794	52	0.90	12.12	56.33 (56.027, 56.641)
NETHERLANDS	1504	61	99.30	14.60	99.24 (99.187, 99.294)
BRAZIL	1481	46	0.00	9.42	0.00 (0.000, 0.000)
AUSTRALIA	1373	49	6.80	11.69	35.23 (34.934, 35.530)
ISRAEL	1330	56	94.90	15.19	99.62 (99.581, 99.657)
SWEDEN	1250	52	68.10	15.72	99.82 (99.793, 99.846)
UKRAINE	1250	29	0.00	4.97	0.00 (0.000, 0.000)
TAIWAN	1160	39	0.00	7.74	0.00 (0.000, 0.000)
BELGIUM	933	48	82.80	14.01	94.72 (94.585, 94.863)
AUSTRIA	751	47	98.00	16.00	99.27 (99.218, 99.324)

**Table 1.** Cont.

Country	Articles ( <i>n</i> )	<i>h</i> -index		Mean citation	
		<i>h</i> -index	Percentile ( <i>q<sub>n</sub></i> )	Mean citation	Percentile ( <i>q<sub>n</sub></i> )
DENMARK	746	55	100.00	19.58	99.99 (99.981, 99.994)
MEXICO	692	29	0.00	7.65	0.00 (0.000, 0.000)

Papers published in 1998 and their citations until 2007.

For each country the merit index  $q_n$  is obtained using as  $W$  the total set of papers published in the World in Physics in 1998 (45718 papers) and their citations received until 2007. For the  $h$ -index the value  $q_n$  is computed analytically and for the mean citation  $q_n$  is approximated using the simulation approach described in the paper with 100,000 random subsets. The corresponding 95% binomial confidence intervals appear in brackets. The figures have been rounded up to three decimal places.

doi:10.1371/journal.pone.0049156.t001

probability that we obtain by chance articles with such a high mean citation, or  $h$ -index, from the set of world papers must be close to zero. As before, this lack of discrimination among top departments can be fixed by considering a different population of interest  $W^*$ . In addition, for the case of the mean citation we can increase the number of random subsets used to estimate  $q_n(x)$ . So far, in our empirical results we have always drawn 100,000

**Table 2.** Field of Clinical Medicine.

(1) Country	(2) Articles <i>n</i>	(3) <i>h</i> -index	<i>W</i> = 155178 <i>W</i> *= 119390	
			(4) Percentile ( <i>q<sub>n</sub></i> )	(5) Percentile ( <i>q<sub>n</sub></i> )
USA	56463	284	100.00	100.00
GERMANY	13822	144	16.68	0.00
ENGLAND	13243	162	99.99	92.03
FRANCE	9556	140	99.45	55.40
ITALY	7471	140	100.00	99.89
CANADA	6297	143	100.00	100.00
NETHERLANDS	4789	123	100.00	99.96
AUSTRALIA	4081	107	99.53	71.14
SWEDEN	4030	114	100.00	99.56
SWITZERLAND	3080	105	100.00	99.80
BELGIUM	2470	94	99.98	96.82
SCOTLAND	2016	90	100.00	99.47
FINLAND	1946	92	100.00	99.97
DENMARK	1841	86	99.99	98.43
NORWAY	1187	71	99.76	90.80

Papers published in 1998 and their citations until 2007.

Column (4): the merit index  $q_n$  is obtained using as  $W$  the total set of papers published in the World in Clinical Medicine in 1998 (155.178 papers) and their citations received until 2007. Column (5): the merit index  $q_n$  is obtained using as  $W^*$  the total set of papers published in these 15 countries in Clinical Medicine in 1998 (119.390 papers) and their citations received until 2007. The figures have been rounded up to three decimal places.

doi:10.1371/journal.pone.0049156.t002

**Table 3.** Sub-field of Neurosciences.

		<i>h</i> -index		Mean citation	
(1)	(2)	(3)	(4)	(5)	(6)
Institution	Articles	<i>h</i> -index	Percentile	Mean	Percentile
	( <i>n</i> )		( <i>q<sub>n</sub></i> )	citation	( <i>q<sub>n</sub></i> )
Yale	209	53	100.00	48.01	100.00
					(100.00, 100.00)
Massachusetts Gen Hosp	186	62	100.00	69.26	100.00
					(100.00, 100.00)
Howard Hughes Med Inst	172	76	100.00	90.58	100.00
					(100.00, 100.00)
Stanford University	133	43	100.00	55.29	100.00
					(100.00, 100.00)
Rockefeller University	73	35	100.00	49.32	100.00
					(100.00, 100.00)
MIT	64	31	99.99	57.05	100.00
					(100.00, 100.00)
Salk Inst Biol Studies	59	34	100.00	78.34	100.00
					(100.00, 100.00)
Brigham & Womens Hosp	44	23	99.77	39.09	98.13
					(98.044, 98.212)
National Institute of Aging (NIA)	40	19	89.50	32.95	89.82
					(89.637, 90.011)
Amgen	29	19	99.89	62.28	99.96
					(99.946, 99.971)
Smithkline Beecham Pharmaceut	24	17	99.86	52.75	99.49
					(99.442, 99.531)
Rush Presbyterian St Lukes Med	24	14	86.79	32.33	84.05
					(83.821, 84.275)
University Fribourg	20	13	92.49	46.55	98.06
					(97.971, 98.141)
Princeton	20	14	98.32	43.05	96.64
					(96.523, 96.747))
Beth Israel Med Ctr	19	12	84.81	40.74	94.94
					(94.804, 95.076)
Natl Inst Med Res	18	12	90.67	46.94	97.66
					(97.569, 97.757)
Mayo Clin Jacksonville	16	14	99.98	43.56	95.56
					(95.431, 95.687)
Max Delbruck Ctr Mol Med	16	13	99.69	34.38	85.51
					(85.296, 85.732)
Cold Spring Harbor Lab	12	10	98.14	56.33	98.44
					(98.366, 98.520)

Papers published in 1998 and their citations until 2007.

For each institution the merit index  $q_n$  is obtained taking  $W$  as the total set of papers published in the World in Neuroscience in 1998 (21876 papers) and their citations received until 2007. For the  $h$ -index the value  $q_n$  is computed analytically and for the mean citation each  $q_n$  is approximated using the simulation approach described in the paper with 100,000 random subsets. The corresponding 95% binomial confidence intervals appear in brackets. The figures have been rounded up to three decimal places.

doi:10.1371/journal.pone.0049156.t003

**Table 4.** Sub-field of Economics.

		<i>h</i> -index		Mean citation	
(1)	(2)	(3)	(4)	(5)	(6)
Institution	Articles	<i>h</i> -index	Percentile	Mean	Percentile
	( <i>n</i> )		( <i>q<sub>n</sub></i> )	citation	( <i>q<sub>n</sub></i> )
Chicago Univ	77	27	100.00	45.90	100.00
					(100.00, 100.00)
Berkeley Univ	70	17	99.82	20.27	99.82
					(99.792, 99.844)
Penn Univ	67	23	100.00	21.70	99.89
					(99.871, 99.912)
Northwestern Univ	65	17	99.92	17.37	99.15
					(99.090, 99.204)
MIT	62	31	100.00	38.73	100.00
					(100.00, 100.00)
Univ Maryland	62	15	98.71	14.21	97.72
					(97.628, 97.813)
Stanford	57	21	100.00	24.12	99.92
					(99.899, 99.953)
Univ Minnesota	42	14	99.87	13.40	94.43
					(94.287, 94.571)
Princeton	36	18	100.00	34.61	99.99
					(99.9777, 99.992)
Duke Univ	28	12	99.94	16.46	97.07
					(96.963, 97.172)
Univ Virginia	26	9	92.45	17.85	98.00
					(97.909, 98.083)
Univ Carlos III Madrid	26	7	47.40	5.84	21.39
					(21.130, 21.639)
Boston Univ	22	13	99.99	34.27	99.69
					(99.651, 99.721)
Univ Iowa	22	8	89.53	9.95	72.47
					(72.197, 72.751)
Boston Col	20	10	99.86	9.10	65.69
					(65.400, 65.988)
Univ Oklahoma	18	6	61.28	5.17	19.08
					(18.833, 19.321)
Univ Pompeu Fabra	17	8	98.11	11.88	83.07
					(82.836, 83.300)
Univ Texas Austin	15	9	99.94	17.87	95.71
					(95.584, 95.836)
Insead	14	8	99.65	24.79	98.82
					(98.753, 98.887)
Univ Miami	10	5	87.45	11.50	79.95
					(79.7000, 80.196)

Papers published in 1998 and their citations until 2007.

For each institution the merit index  $q_n$  is obtained taking  $W$  as the total set of papers published in the World in Economics in 1998 (7542 papers) and their citations received until 2007. For the  $h$ -index the value  $q_n$  is computed analytically and for the mean citation each  $q_n$  is approximated using the simulation approach described in the paper with 100,000 random subsets. The corresponding 95% binomial confidence intervals appear in brackets. The figures have been rounded up to three decimal places.

doi:10.1371/journal.pone.0049156.t004



random subsets (for each  $n$ ). This might be more than enough for intermediate percentiles but not for percentiles very close to 100. However, given the purpose of this paper, we find of no practical importance that, for example, the differences in Table 3 between percentiles 99.96 and 99.49 are not statistically significant.

## Discussion

In this paper we have proposed a new simple and intuitive method to assess the citation merit of any set of scientific papers in any field. One advantage of our approach is that it can be applied to a variety of problems. For example, it might be applied to rank scientific journals. The merit of a given journal that publishes  $n$  articles in a year in a given field would be given by the probability that a subset of  $n$  articles in that field are of lower quality according to some criterion as the mean citation or the  $h$ -index (note that the merit of a journal is not the same as the merit of the authors who publish in the journal). A second advantage is the possibility of comparisons of the scientific merit of research units in different fields. This can be done because the merit of each research unit is associated with a probability (or percentile) that might be reasonable to compare across different fields.

As far as the international cooperation is concerned, it is well known that domestic and international publications are characterized by very different citation rates. Therefore, using whole counts as we have done in this paper, or following [20] recommendation in favor of using fractionalized counts to calculate citation indicators at the national level, might make a significant difference that it would be convenient to investigate.

In the empirical application of the method we have used two well-known and vastly different citation impact indicators: the mean citation and the  $h$ -index. However, recall that, given their high skewness, the upper and lower parts of citation distributions are typically very different. Consequently, average-based indicators -such as the mean citation- may not adequately summarize these distributions. On the other hand, both the  $h$ -index and many

of the indicators of the same family have been criticized by some authors (see [25–27]). Therefore, it may be worthwhile to study the merit of research units according to some of the new indicators that are rapidly being suggested (see [26,28–32]).

It is important to note that our approach is not trying to make any inference on the underlying model explaining the scientific output of the different units. For an overall assessment of the relative merit or performance of a research unit we should take into account many other variables, such as the budget, number of researchers, etc. Two research units with the same merit according to a set of citation indicators as understood in this paper may vastly differ in the productivity of its research staff or, more generally, in the efficiency with which scientific results are obtained from a complex input vector. Thus, we only provide a method to assess a research unit's performance in a certain dimension, quite independently of the underlying model explaining why different units produce scientific publications of different citation impact and citation merit.

## Supporting Information

**Tables S1 Analysis of the 40 countries with the largest number of papers in the corresponding field.** (XLSX)

## Acknowledgments

Conversations with P. Albarrán, J. Rodríguez-Puerta and P. Poncela are gratefully acknowledged. We also thank M.D. Collado and two anonymous referees for their helpful comments.

## Author Contributions

Conceived and designed the experiments: JAC IO-O JR-C. Performed the experiments: JAC IO-O JR-C. Analyzed the data: JAC IO-O JR-C. Contributed reagents/materials/analysis tools: JAC IO-O JR-C. Wrote the paper: JAC IO-O JR-C.

## References

- Molinari J, Molinari A (2008) A new methodology for ranking scientific institutions. *Scientometrics* 75: 163–174.
- Kinney A (2007) National scientific facilities and their science impact on nonbiomedical research. *PNAS* 104: 17943–17947.
- Moed H, Burger W, Frankfort J, van Raan A (1985) The use of bibliometric data for the measurement of university research performance. *Research Policy* 14: 131–149.
- Moed H, Brui RD, van Leeuwen T (1995) New bibliometrics tools for the assessment of national research performance: Database description, overview of indicators, and first applications. *Scientometrics* 33: 381–422.
- Van Raan AFJ (2004) Measuring Science. *Handbook of Quantitative Science and Technology Research*, 19–50 Moed et al. (eds.).
- Schubert A, Glänzel W, Braun T (1983) Relative citation rate: A new indicator for measuring the impact of publications. *Proceedings of the First National Conference with International Participation in Scientometrics and Linguistics of Scientific Text*. Varna.
- Braun WG, Schubert A (1985) *Scientometrics indicators. A 32 country comparison of publication productivity and citation impact*. Singapore, Philadelphia: World Scientific Publishing Co. Pte. Ltd.
- Schubert A, Braun T (1986) Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics* 9: 281–291.
- Glänzel W, Schubert A, Braun T (2002) A relational charting approach to the world of basic research in twelve science fields at the end of the second millennium. *Scientometrics* 55: 335–348.
- Vinkler P (1986) Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics* 10: 157–177.
- Vinkler P (2003) Relations of relative scientometric indicators. *Scientometrics* 58: 687–694.
- Fortunato S, Radicchi F, Castellano C (2008) Universality of citation distribution: Toward an objective measure of scientific impact. *PNAS* 105: 17268–17272.
- Bornmann L, Leydesdorff L (In press) Which cities produce more excellent papers than can be expected? a new mapping approach -using google maps- based on statistical significance testing. *Journal of the American Society for Information Science and Technology*.
- Hirsch J (2005) An index to quantify an individual's scientific research output. *PNAS* 102: 16569–16572.
- Albarrán P, Crespo J, Ortuño I, Ruiz-Castillo J (2011) The skewness of science in 219 subfields and a number of aggregates. *Scientometrics* 88: 385–397.
- Seglen P (2004) The skewness of science. *Journal of the American Society for Information Science* 43: 628–638.
- Schubert A, Glänzel W, Braun T (1985) A new methodology for ranking scientific institutions. *Scientometrics* 12: 267–292.
- Glänzel W (2007) Characteristic scores and scales: A bibliometric analysis of subject characteristics based on long-term citation observation. *Journal of Informetrics* 1: 92–102.
- Albarrán P, Ruiz-Castillo J (2011) References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology* 62: 40–49.
- Aksnes D, Schneider J, Gunnarsson M (2012) Ranking national research systems by citation indicators. a comparative analysis using whole and fractionalised counting methods. *Journal of Informetrics* 6: 36–43.
- May R (1997) The scientific wealth of nations. *Science* 275: 793–796.
- King D (2004) The scientific impact of nations. *Nature* 430: 311–316.
- Albarrán P, Crespo J, Ortuño I, Ruiz-Castillo J (2010) A comparison of the scientific performance of the u.s. and europe at the turn of the 21st century. *Scientometrics* 88: 329–344.
- Herranz N, Ruiz-Castillo J (2012) Multiplicative and fractional strategies when journals are assigned to several sub-fields. *Journal of the American Society for Information Science and Technology*. In press.
- Bouyssou D, Marchant T (2011) Ranking scientists and departments in a consistent manner. *Journal of the American Society for Information Science and Technology* 62: 1761–1769.
- Bouyssou D, Marchant T (2011) Biobibliometric rankings of journals based on impact factors: An axiomatic approach. *Journal of Informetrics* 5: 75–86.



27. Waltman L, van Eck N (2012) The inconsistency of the *h*-index. *Journal of the American Society for Information Science and Technology* DOI: 10.1002/asi.21678. In press.
28. Albarrán P, Ortuño I, Ruiz-Castillo J (2011) High- and low-impact citation measures: empirical applications. *Journal of Informetrics* 5: 122–145.
29. Ravallion M, Wagstaff A (2011) On measuring scholarly inuence by citations. *Scientometrics*, 88: 321–337
30. Leydesdorff L, Bornmann L (2011) Integrated impact indicators (i3) compared with impact factors (if): An alternative research design with policy implications. *Journal of the American Society for Information Science and Technology* 62: 2133–2146.
31. Leydesdorff L, Bornmann L, Mutz R, Ophof T (2011) Turning the tables on citation analysis one more time: Principles for comparing sets of documents. *Journal of the American Society for Information Science and Technology* 62: 1370–1381.
32. Rousseau R (2012) Basic properties of both percentile rank scores and the i3 indicator. *Journal of the American Society for Information Science and Technology* DOI: 10.1002/asi.21684. In press.