Cold Spring Harbor Laboratory
bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

# *fCite*: a fractional citation tool to quantify an individual's scientific research output

Lukasz Pawel Kozlowski[1] *

1 Solid Scientometrics Sp. z o.o., Bieliny 26-004, Holy Cross Province, Poland
* lukasz.kozlowski.lpk@gmail.com

SHORT TITLE: **fCite: dividing impact by the number of authors**

**ABSTRACT:** Here, I present the *fCite* web service (fcite.org) a tool for the in-depth analysis of an individual's scientific research output. While multiple existing tools (e.g., Google Scholar, iCite, Microsoft Academic) focus on the total number of citations and the H-index, I propose the analysis of the research output by considering multiple metrics to provide greater insight into a scientist's multifaceted profile. The most distinguishing feature of *fCite* is its ability to calculate fractional scores for most of the metrics currently in use. Thanks to the division of citations (and RCR scores) by the number of authors, the tool provides a more detailed analysis of a scholar's portfolio. *fCite* is based on PUBMED data (~18 million publications), and the statistics are calculated with respect to ORCID data (~600,000 user profiles).

**ABBREVIATIONS:** RCR – Relative Citation Ratio; PMID – **P**ub**M**ed **Id**entifier; ORCID – **O**pen **R**esearcher & **C**ontributor **ID;** DORA - San Francisco **D**eclaration **o**n **R**esearch **A**ssessment; FLAE model - **f**irst-**l**ast-**a**uthor-**e**mphasis model; EC model – **e**qual **c**ontribution model; M-index – H-index divided by the number of years; fH-index – before H-index is calculated the citations are divided using FLAE model; fM-index – fH-index divided by the number of years; $FLAE_{RCR}$ – RCR score calculated using FLAE model; $EC_{RCR}$ – RCR score calculated using EC model; $FLAE_{cit}$ – number of citations calculated using FLAE model; $EC_{cit}$ – number of citations calculated using EC model; SJR - **S**CImago **J**ournal **R**ank indicator; HCR – **H**ighly **C**ited **R**esearchers list

**KEY WORDS:** bibliometrics, scientometrics, bibliometric analysis, quality of publications, bibliometrics tools, science impact, science policy, science evaluation, research assessment

## INTRODUCTION

At present, the impact of the work of a scientist can be estimated by a number of bibliometric metrics, but there is a strong bias towards the number of articles written by an author, the total number of citations of those articles, the impact factors of the journals in which they appeared (1) and finally the H-index (2). In contrast to this approach, increasing number of people are opposed to a bibliometric, mechanical *modus operandi* and in favour of expert assessment (e.g., see the DORA declaration) (3). Although expert approach is a compelling idea, in real life a fair assessment of a scientific portfolio comprising multiple publications (for instance containing over 100 items) spread across multiple journals (e.g., in 2019 PUBMED alone indexed 48,601 journals) may not be possible in a reasonable amount of time. Even if the expert is familiar with the quality of the journal in which the publications appeared and, even if he/she has read the most important publications from the portfolio, he/she still needs to understand and judge the author's contribution to given work(s). With multiple-author papers this can be very difficult to achieve. Acknowledgment statements (if any) are usually very frugal, and it is often impossible to say which part of the work was done by which author. Moreover, if one also recognizes that publications are frequently interdisciplinary, the proper assessment of the influence of the average-sized portfolio is beyond the scope of a single person, and ultimately, it is very subjective. On the other hand, in many fields the order in the author list can be considered a rough approximation of the contribution, where the first author is the scientist who performed the most of the experiments (e.g., a PhD student), the middle authors are those who helped with multiple specialized parts of the work or/and the analysis of the data and the last/corresponding author is a principal investigator who conceived the project, obtained the funding and supervised all steps (frequently this does not exclude involvement in the experiments or the analysis). Such a model is termed the **f**irst-**l**ast-**a**uthor-**e**mphasis (FLAE) model (4), and depending on how much emphasis is placed on a particular author, the FLAE model can have multiple flavours (here we use three models named FLAE, FLEA2, and FLAE3; for details see the *Material and Methods*). However, given a sufficient number of items in the portfolio, they yield very similar results (5). In contrast, if the order of authors is random or alphabetical, we can always use the equal contribution (EC) model in which each author has the same weight (**Fig. 1**, **Supplementary Fig. 1-4, Supplementary Tables 1-5**). At this point, an open question is how to assess the influence of the publication. The most accessible (and frequently used) metric is the number of citations it has received over time. Usually, the metrics such as total citation counts or the H-index are calculated using global scores (regardless of the number of authors, each author obtains all of the citations of the publication), but applying FLAE or EC models provides a straightforward way to quantify the author's contribution in a more precise manner. The division of the contribution is a highly demanding (and overlooked (6)) feature because the number of authors has increased steadily over time to exceed an average of six authors per research publication in 2015 (**Fig. 2**, **Supplementary Table 6)**. The trend of having increasing numbers of authors is also clear when we analyse the mode of the number of authors in publications over the last 25 years (**Fig. 3**, **Supplementary Table 7**). Currently, publications with hundreds of authors are not rare, and some items can have more than several thousand authors. Concomitantly, shared first or last authorship has become a common practice, and it is not difficult to find publications with three or more shared first authors and few corresponding/last authors.

Rewarding all authors, regardless of their number, is an obvious shortcoming of current bibliometric tools and contradicts Price's law of productivity according to which the square root of the number of people is responsible for 50% of the work. This means that in 10-author publication, 3 of them do ½ of work (7). The remaining work is done by 7. Similarly, out of 100 authors, around 10 do the half of work and 90 second half.

Consider a hypothetical situation in which you are a member of a grant or fellowship committee and have two applicants. Both of them published single publications in the same journal (for simplicity of the example), but the first publication has two authors (the applicant and his/her supervisor), while in the second case the applicant is the middle author of a consortium paper (such papers usually have few hundred authors). After a few years since the publication date, you see that the first publication has received a few dozen citations, while the second has a few hundred citations. Which candidate would you prefer? In the presented example, most experienced assessors would prefer the first candidate. If you were to do some back-of-the-envelope calculations, you would conclude that the first item has roughly an order of magnitude more citations per author than the second, and it is almost trivial to assess the contribution to the first publication; however, when there are a few hundred authors, it is literally impossible to say who did what, and most likely those hundreds of citations are self-citations or/and courtesy citations (for instance, the publication describes an important resource used by the whole field). The presented example is highly simplified, and usually there are more items in portfolios, which significantly complicates the analysis. Typically, portfolios will be more diverse in the number of items and journals in which they were published. The other well known drawback of the expert based evaluation system is it's high cost (e.g, Research Excellence Framework in United Kingdom) (8). The *fCite* web service presented here should fill the gap between the (overly) simple bibliometric and expensive expert-based approaches, facilitating fairer assessment of scientific output.

## MATERIALS AND METHODS

Data sets

The statistics in fCite are based on two data sets: (1) PUBMED data set – contains PUBMED publications (17,787,016 publications; 14,444,982 research and 3,342,034 non-research items) obtained via icite.od.nih.gov portal (**Supplementary Data 1-2**) and (2) ORCID (**O**pen **R**esearcher & **C**ontributor **ID**) data set – contains 5,380,983 user profiles (**Supplementary Data 3**).

Fractional contribution models of the authorship

Four, different models had been used to assess the author contribution:

- FLAE (**f**irst-**l**ast-**a**uthor-**e**mphasis) model is based on Tscharntke et al. 2007 definition with slight modifications (4). The contribution of individual authors can be described briefly as "the first author gets 100, the last 50, and all others 100/number of authors and then scores are normalized to 1". This type of the model gives the strongest weights to the first and the last author penalizing middle authorship (**Supplementary Table 1**).
- FLAE2 model is based on Corrêa Jr. et al. 2017 (5). This is empirical model based on the authorship contribution for the mega-journal PLoS ONE (~65,000 publications). On average this model is more benign for middle authors. As the data presented by Corrêa Jr. and co-authors are limited up to ten authors, for the longer author lists the contribution has been modeled by curve fitting with some noise using the initial matrix (with up to ten authors), and thresholds 0.06 for 30 authors and 0.07 for 100 authors (for more details see scipy.optimize.curve_fit documentation and http://www.fcite.org/FLAE2.txt) (**Supplementary Table 2**). Additionally, this model is asymmetric, i.e. middle author weights depends on the position, the closer to the first author, the better are weights for middle author (up to $10^{th}$ author).

- FLEA3 model is a simple variation of FLAE model, but the contribution of individual authors is more equal. It can be shortly described as "the first author always get at least three times more than co-authors, and the last author at least two times more than other co-authors" (**Supplementary Table 3**).
- EC (equal contribution) model assumes that each author contributed equally to given work (**Supplementary Table 4**).

First three models assume that for a given field the order of the authors is not random and the first author was the one who contributed the most while the last is a senior author who conceived the project (frequently the corresponding author). Such assumption is true for many sub-fields of the biomedical sciences. Alternatively, in many other sub-fields the order of the authors can be alphabetical, ordered from the most significant to the least significant author or completely random or irregular. In such cases EC model should be used. For simplicity, all models assume that there is only one first and one last author which is not necessarily true as with strong pressure for the publishing in the top journals and having more and more authors per the work, nowadays many papers have multiple first and senior authors. All four models are metrics agnostics thus they can be used for the citations, RCR scores, or/and H-index.

Additional metrics used in *fCite*

In order to analyze the portfolio, the user is asked to provide all combination of author names and surnames, and the list of PMID (**P**ub**M**ed **ID**entifier) ids . As a result he/she obtains:

a) the size of the portfolio with the time span of the publishing period,
b) the number (and the percentage) of the single, the first, the last and the middle author papers,
c) H-index (2),
d) M-index (H-index divided by the number of years from the first publication),
e) fH-index and fM-index (the citations are divided according the author contribution to each paper using FLEA model),
f) the average number of the papers per year,
g) the total and fractional citation and RCR scores based on FLAE, FLAE2, FLAE3, EC models (Citations, total RCR, $FLAE_{RCR}$, $FLAE2_{RCR}$, $FLAE3_{RCR}$, $EC_{RCR}$, $FLAE_{cit}$, $FLAE2_{cit}$, $FLAE3_{cit}$, $EC_{cit}$, respectively),
h) the average number of the authors,
i) FLEA per year (RCR),
j) the average FLAE article score (RCR),
k) the average article impact per year (RCR),
l) the ratio between $FLAE_{RCR}$ and total RCR and the expected value,
m) sortable table for individual publications with PMID, year, title, authors, article-type, journal, and $FLAE_{RCR}$, $FLAE2_{RCR}$, $FLAE3_{RCR}$, $EC_{RCR}$, $FLAE_{cit}$, $FLAE2_{cit}$, $FLAE3_{cit}$, $Ec_{cit}$, Citations, total RCR scores.

Initial data cleaning

In order to analyze the authorship patterns, the PUBMED data set (over 17 million of publications) had been mapped into ORCID portfolios (over 5 million of users). The ORCID data set provided author name and surname with the list of publications. First, an empty records (the profiles without public data) had been discarded (4,217,452 out of 5,380,983 records). Next, the portfolios with at least one publication with the DOI, PMID or PMC identifiers had been filtered. This gave 1,154,443 portfolios (with 19,516,285 non-unique articles in total). As 19,097,891 (97,85%) of items had only DOI identifier, additional step was required (namely mapping DOI to PMID identifiers). The whole PUBMED records (27,414,004 publications) had been

search for DOI using summary XML files and *eutils* tool provided by the National Institutes of Health (NIH). As a result 599,468 (7,813,971 articles) of non-empty portfolios with at least one PMID had been obtained.

Example record from ORCID (csv format)

```
ORCID,surname,name,list_of_PMIDS
0000-0002-2518-5940,Liebovitz,David,23550982||23646091||19468082||22034582||19267397||17219478||
19647184||28527507||17219519
```

Example record from PUBMED (json format)


```
{
    "pmid": 23456789,
    "doi": "10.1002/cncr.27976",
    "authors": "Arun Sharma, Stephen M Schwartz, Eduardo M\u00e9ndez",
    "citation_count": 26,
    "citations_per_year": 4.333333,
    "expected_citations_per_year": 2.538138,
    "field_citation_rate": 4.872565,
    "is_research_article": true,
    "journal": "Cancer",
    "nih_percentile": 69.700000,
    "relative_citation_ratio": 1.707288,
    "title": "Hospital volume is associated with survival but not  multimodality therapy in Medicare
patients with advanced head and neck cancer.",
    "year": 2013
}
```


*fCite* uses following fields: authors, citation_count, relative_citation_ratio, is_research_article, year and pmid.

Data analysis

One of the first steps of the analysis was to clean the name and surname provided by the ORCID database. The data in ORCID are in the UNICODE (UTF-8) format which means that they can contain any Non-English letters. Thus, at this step all surnames and names had been translated to equivalents of English letters (e.g., Kozłowski Łukasz to Kozlowski Lukasz, 吴锋 to Wu Feng). Then, given the list of PMIDs in the portfolio, all publication records from PUBMED had been retrieved. In order to identify the author position on the authorship list, Levenshtein and Jaro–Winkler distances had been applied in the following way. First, a set of possible surname and name combinations had been prepared (name surname, surname name, n surname, name initial surname, etc. for instance given John Smith, the set contained john smith, smith john, j smith, john x smith, etc.). This step was required as the order of the name, surname, initials and the letter size are frequently different in the databases or/and particular publication records. Next, for each author in the individual authorship list the Jaro–Winkler distance is calculated. The author which has the highest Jaro-Winkler distance is used (the similarity threshold of 0.7 is used to filter out non-important hits). From now on, for each publication in the portfolio, the position of the author is available and can be used to divide the publications into the sole, first, last and middle author ones. Having positions of the authors allow to use fractional models (FLAE, FLAE2, FLAE3 and EC) to calculate fractional scores for the citations and RCR metrics.

ORCID data had been also used to quantify the significance of obtained scores. For instance, it is not enough to say that $FLAE_{RCR}$ (or any other score) is equal to 10. Obviously, the bigger the number the better, but it is useful to compare it to some reference. For this purpose we calculated the percentiles for the score in respect to all

ORCID portfolios (the value below which a given percentage of observations in a group of observations falls). Note that the percentiles presented by *fCite* are calculated separately for each individual metrics including division into research and non-research item portfolios (**Table 1**). Additionally, as the distribution of bibliometric metrics is practically never normal (Gaussian) (**Supplementary Fig. 5**), the percentiles are presented with additional precision (this allow to distinguish similar portfolios, especially top ones).

## RESULTS

The primary result of the work presented here is, the web service fcite.org , where an author's contribution can be calculated using fractional models in addition to the plethora of statistics related to a given portfolio (**Fig. 4**). The *fCite* service operates using a list of PMID and/or ORCID ids accompanied by all combinations of the names and surnames of a given author. The analysis can be performed for all items (research articles and non-research items such as editorials, reviews, and others) or separately only for the research items. The *fCite* service relies on citations and so-called RCR (**r**elative **c**itation **r**atio) scores that come from the iCite web service and are calculated based on the PUBMED database (9). As of October 2019, *fCite* comprises over 17 million publications and counting (the *fCite* database grows by ~100,000 items each month). The reference for the scores comes from the analysis of ORCID data (572,910 profiles with 7,008,012 unique publications in total). The ORCID profiles provide the names of the authors, together with the lists of co-authored publications. Thus, by using string metrics such as Levenshtein and Jaro–Winkler distances (10), it is possible to identify an author's position in the list of the authors for each publication. This provides the unique opportunity to study authorship patterns depending on whether the person is the first, middle, last or single author. These data provide a solid foundation for the assessment of the author's position importance with respect to portfolio size and time. As shown in **Fig. 5** (and **Supplementary Tables 8-9**), at the beginning of a scientist's career (with small portfolios with fewer than 10 items), a substantial number of first author papers are expected. As the researcher progresses (and the number of publications increases), the last author publications begin to take the place of the first author publications. Surprisingly, the middle and single authored fractions are roughly stable regardless of portfolio size (where single author papers are extremely rare, and middle author papers constitute over half of the items). Moreover, calculating the main scores provided by *fCite* (e.g., $FLAE_{RCR}$, $FLAE2_{RCR,}$ $FLAE_{Cit}$, $FLAE2_{Cit}$, $EC_{RCR}$, $EC_{Cit,}$) for the ORCID users provides solid ground for the assessment of the importance of the obtained numbers (so-called percentiles) (**Table 1**).

## DISCUSSION

Over the past fifty years, bibliometrics have become an inherent part of the assessment of scientific progress. Such measures, with all of their pros and cons, will be used regardless of whether we support this approach. This development began with the impact factors defined by Eugene Garfield in the 1950s and peaked with the creation of the H-index in 2005. Despite their multiple shortcomings, bibliometrics can offer an instantaneous and relatively fair assessment of science impact. As many scholars have noted previously, one cannot focus on a single number because it cannot embrace the complexity of a researcher's work (see also Goodhart's adage). Therefore, it is not proper to focus, for instance, only on the total number of citations or the H-index (which is itself highly correlated with the total number of citations (11)). As those two metrics have been frequently used by funding bodies (consciously/openly or not), researchers have optimized their behaviour, which has led to citation cartels (12), salami slicing (13), and continual increases in the number of authors per publication and the self-citation rate (14). A so-called "publish-or-perish" culture has emerged. The purported quality of a work

is inherited directly from impact factors of the journal immediately after publication. The number of authors of papers is irrelevant because all of them receive full credit. Even if someone were to state that the first or the last/ corresponding authors are more important, the community has already found easy "fix" by adding multiple first and last authors. This may sound pessimistic, but some fields have already adapted to this new reality very well. Therefore, no one is surprised when a high-energy physics paper has a few thousand authors. Similar approaches are emerging in other fields. In medicine, which already has one of the highest number of authors per publication, many believe that the data provider should be listed as an author of all subsequent publications even without making any other contribution (15). Recently, there is a growing trend towards establishing consortia or groups containing multiple labs/consortia. While this has the advantage of making collaborations that can have synergistic effects, it also has disadvantages. Usually, such initiatives are based on multi-million-dollar grants, and as the results appear, the whole group (usually a few hundred authors) is assigned as the authors of almost every paper produced by the consortium. The *fCite* proposed here stymies such malicious behaviour by simply dividing the citations (or RCRs) by the number of authors while taking into account an author's number and position. Obviously, the FLAE models used here are far from perfect, and they cannot replace an experienced assessor who reads publications and is familiar with all the insights of his/her own field, but they are a good starting point and certainly a better solution than using the total number of citations or H-index.

All of this being said, one should be aware of the multiple limitations of *fCite*. First, as *fCite* is based on PUBMED, it is not appropriate for many fields that are not well represented in that database (for instance, computer science or social sciences). Second, *fCite* currently does not filter out self-citations (under development). Moreover, the author has this far been unable to tag and fix all consortium/group papers. Frequently, such items' authorship appears as "John Smith, Jan Kowalski, SOME Group", where the first two are leaders, and the SOME group consists of a hundred or more people whose names appear in the supplementary material. FLAE models will count such cases as papers with three authors (likely to have hundreds of citations designating only those three authors and elevating the scores for the first two). Finally, the last shortcoming of *fCite* is that it accords all citations the same weight, which is a massive simplification. This aspect of bibliometrics is well studied and can be considered from many angles. First, not all citations are equal, as a citation can be positive or negative (where the subsequent authors disagree with the original hypothesis). Then, even if the citation is positive, it may have different meanings depending on the section of the article where it is made (the introduction, the methods or the discussion). Moreover, some citations are more important because they are cornerstones for subsequent research, while others are simply review mentions used briefly in the introduction. The other aspect related to citations is that frequently the most important citations are missing due to journal restrictions (for instance, the entire method section, pivotal for any research, is placed in the supplementary material, which has its own reference list and is not listed in most databases). Many such cases can be handled by semantic methods (16), but this approach remains in its infancy. Other characteristics that are frequently used (but not implemented in *fCite*) are the importance of the journal from which the citation comes (e.g., SJR indicators developed by SCImago (17)). Nevertheless, *fCite* is the first, large-scale method that takes into account the number of authors and their positions (only one-time analyses of specific journals, fields, or nations have been done in the past (18) (19)). Additionally, *fCite* uses RCR scores that are taken from iCite (based on PUBMED). Note that these scores differ from citations in many aspects. First, RCR is intended to capture field relevance (it is normalized with respect to the field's citation levels). Next, in contrast to citations, which are only additive metrics, RCRs can decrease over time. This aspect, while it has been

criticized by some (20), is a very useful and demanded feature of bibliometric metrics (that is also missing in the H-index), as RCR can decline when the work begins to be outdated. The other feature of RCR that should not be overlooked is that this metric gives more weight to newer articles (for instance, ten citations for ten-year-old and two-year-old articles will result in dramatically different RCR scores).This is in aggrement with well recognised observation that the citation over article age is skewed and the most of the citations the publication receives within at the age of 2-5 years (**Supplementary Fig. 6**).

A highly illustrative example is an analysis of top researchers in comparison to the scores provided by Google Scholar and other sources. **Supplementary Table 10** reports a selection of statistics for some successful scientists (many of whom are listed in the Highly Cited Researchers (HCR) list created by Clarivate Analytics). It is clear that the H-index or total citation counts can often be misleading, and more comprehensive analysis using multiple bibliometric metrics can help. For instance, while HCR is most likely filtered against easy-to-spot cases such as Scientists B and D, it still frequently includes cases such as Scientists A and F. On the other hand, due its limitations (having >15 so-called "Highly Cited Papers"), some outstanding scientists are overlooked (e.g., Scientist C). Therefore, *fCite* can be a very useful tool for deep profiling of even very similar portfolios (with respect to the H-index or total citation count) with surprising discriminatory power (e.g., compare Scientists M and N).

In summary, *fCite* (available free of charge at fcite.org) is a bibliometric tool that provides versatile metrics that can take into account the number of authors and their position on the authorship list. Hopefully, it will facilitate unbiased comparisons of researchers' importance when they are competing for limited funding and, consequently, enhance scientific development.

## REFERENCES


1. Garfield, E. (1955) Citation indexes for science; a new dimension in documentation through association of ideas. *Science* **122**, 108–111
2. Hirsch, J. E. (2005) An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16569–16572
3. Schmid, S. L. (2017) Five years post-DORA: promoting best practices for research assessment. *Molecular biology of the cell* **28**, 2941–2944
4. Tscharntke, T., Hochberg, M. E., Rand, T. A., Resh, V. H., and Krauss, J. (2007) Author sequence and credit for contributions in multiauthored publications. *PLoS Biol.* **5**, e18
5. Corrêa Jr., E. A., Silva, F. N., da F. Costa, L., and Amancio, D. R. (2017) Patterns of authors contribution in scientific manuscripts. *Journal of Informetrics* **11**, 498–510
6. Lozano, G. A. (2013) The elephant in the room: multi-authorship and the assessment of individual researchers. *Current Science* **105**, 443–445
7. de Solla Price, D. J. (1965) Little science, big science. Columbia University Press New York
8. Jones, M. M., Manville, C., and Chataway, J. (2017) Learning from the UK's research impact assessment exercise: a case study of a retrospective impact assessment exercise and questions for the future. *The Journal of Technology Transfer* 1–25
9. Hutchins, B. I., Yuan, X., Anderson, J. M., and Santangelo, G. M. (2016) Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level. *PLoS Biol.* **14**, e1002541
10. Cohen, W. W., Ravikumar, P., and Fienberg, S. E. (2003) A Comparison of String Distance Metrics for Name-Matching Tasks.
11. Yong, A. (2014) Critique of Hirsch's citation index: A combinatorial Fermi problem. *Notices of the AMS* **61**, 1040–1050

12. Fister Jr, I., Fister, I., and Perc, M. (2016) Toward the discovery of citation cartels in citation networks. *Frontiers in Physics* **4**, 49
13. Rogers, L. F. (1999) Salami slicing, shotgunning, and the ethics of authorship. *AJR. American journal of roentgenology* **173**, 265–265
14. King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., and West, J. D. (2017) Men set their own cites high: Gender and self-citation across fields and over time. *Socius* **3**, 2378023117738903
15. Longo, D. L. and Drazen, J. M. (2016) Data Sharing. *New England Journal of Medicine* **374**, 276–277
16. Li, D.-C., Liu, H., Chute, C. G., and Jonnalagadda, S. R. (2013) Towards assigning references using semantic, journal and citation relevance.
17. González-Pereira, B., Guerrero-Bote, V. P., and Moya-Anegón, F. (2010) A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of informetrics* **4**, 379–391
18. Abramo, G., D'Angelo, C. A., and Rosati, F. (2013) Measuring institutional research productivity for the life sciences: the importance of accounting for the order of authors in the byline. *Scientometrics* **97**, 779–795
19. Egghe, L. (2008) Mathematical theory of the h-and g-index in case of fractional counting of authorship. *Journal of the American Society for Information Science and Technology* **59**, 1608–1616
20. Janssens, A. C. J., Goodman, M., Powell, K. R., and Gwinn, M. (2017) A critical evaluation of the algorithm behind the Relative Citation Ratio (RCR). *PLoS biology* **15**, e2002536

## ACKNOWLEDGMENTS

The author acknowledge the NCBI, NLM, ORCID and iCite teams. Funding: L.P.K. (Solid Scientometrics). The author declares competing interests. The Solid Scientometrics company owned by L.P.K. had been established to separate and protect the intellectual property of this project. Since November 2018 L.P.K. is employed in the Institute of Informatics, University of Warsaw, where his research is focused on the computational biology (the proteomics). The project and the Solid Scientometrics company had been conceived before L.P.K had been employed in the University of Warsaw.

## AUTHOR CONTRIBUTIONS

L.P.K. conceived the project, acquired and analyzed the data, developed the web service, and prepared the manuscript.

## DATA AND MATERIALS AVAILABILITY

The *fCite* web service (fcite.org) is available free of charge. All the data needed to evaluate the conclusions in the paper are presented in the Supplementary Materials and at the *fCite* web site. The raw data come from NIH (iCite) and ORCID databases and are available as stated in the Supplementary Materials.

**Figure 1.** Fractional models used in *fCite* (FLAE, FLAE2, FLAE3, EC). The weights for the first, middle and last author up to ten authors. For numerical data see **Supplementary Tables 1-3**, respectively.

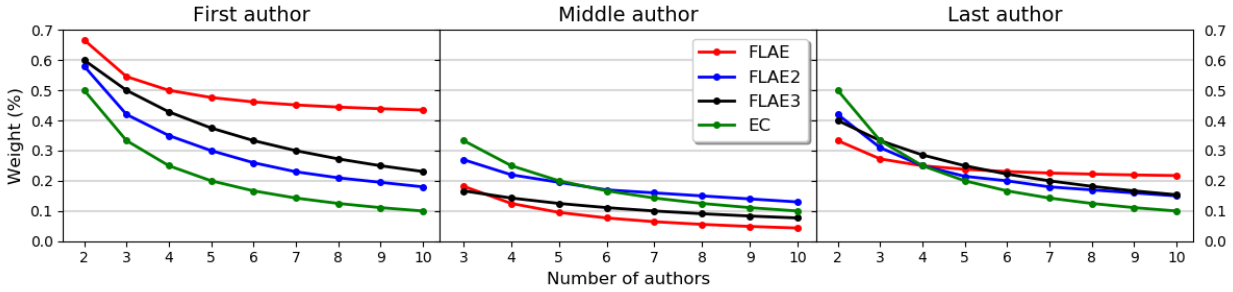



**Figure 2.** The increase in the average number of authors over time (whole PUBMED, 17 million items). For the numerical data, see **Supplementary Table 6**.

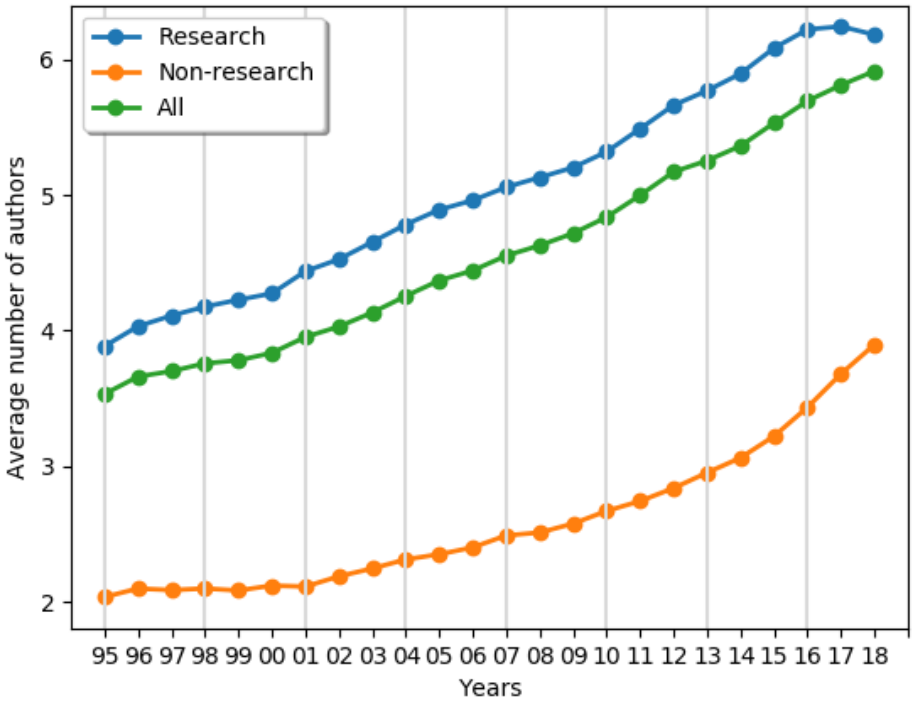


**Figure 3.** Number of authors over time with respect to research and non-research items (the years 1995-2018; 17,651,086 PUBMED publications). For the numerical data see **Supplementary Table 7**. Animated version of the the figure is available at: *http://www.fcite.org/stats.html#Authors_over_years*

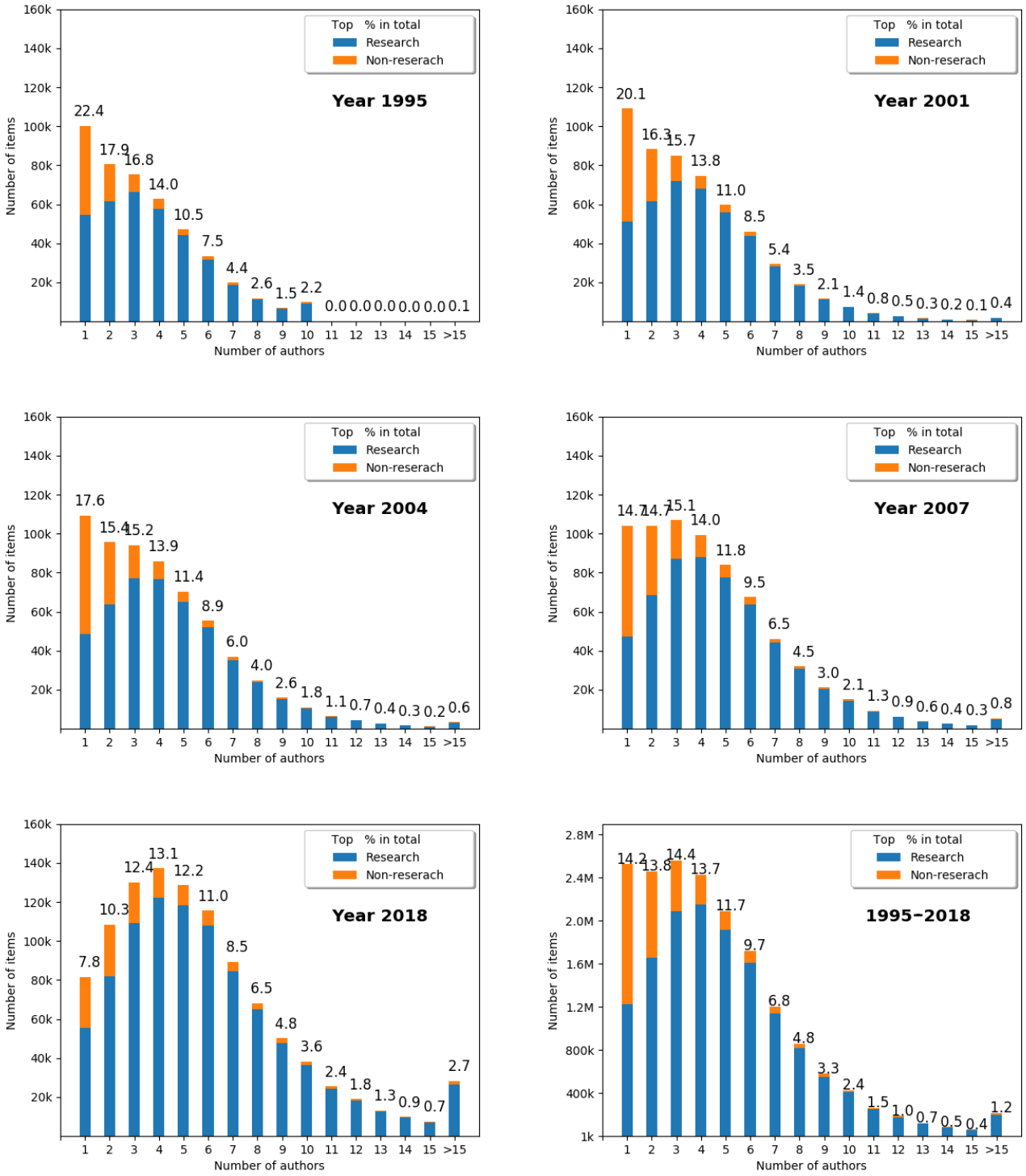


**Figure 4.** Example output from *fCite*. Given the list of PMIDs (16 items in this case) and all combinations of the name, the user obtains a detailed analysis of the researcher's contribution (for a brief user manual for *fCite*, see the "Help" section). On the top left, you have the number of articles and the time span within they were published. Next, in the top-right panel, you have some statistics about single, first, last and middle authorship alongside the H-index, M-index and their fractional analogues. In the centre, you have the scores for the fractional models based on RCR ($FLAE_{RCR}$, $FLAE2_{RCR}$, $FLAE3_{RCR}$, $EC_{RCR}$, $FLAE_{RCR}$ and total RCR). Then, using those values, you obtain scores such as the average number of the authors in the publication, $FLAE_{RCR}$ per year, average article $FLAE_{RCR}$, the article impact per year, and the ratio of $FLAE_{RCR}$/RCR alongside the expected ratio based on the number of authors in the portfolio. The next line is a simple repetition of the scores but here based on raw citations. Note that the main scores are accompanied by the percentiles (based on ORCID portfolios) to facilitate the assessment of the importance of the scores. Finally, you have a sortable table with individual publications and their scores.

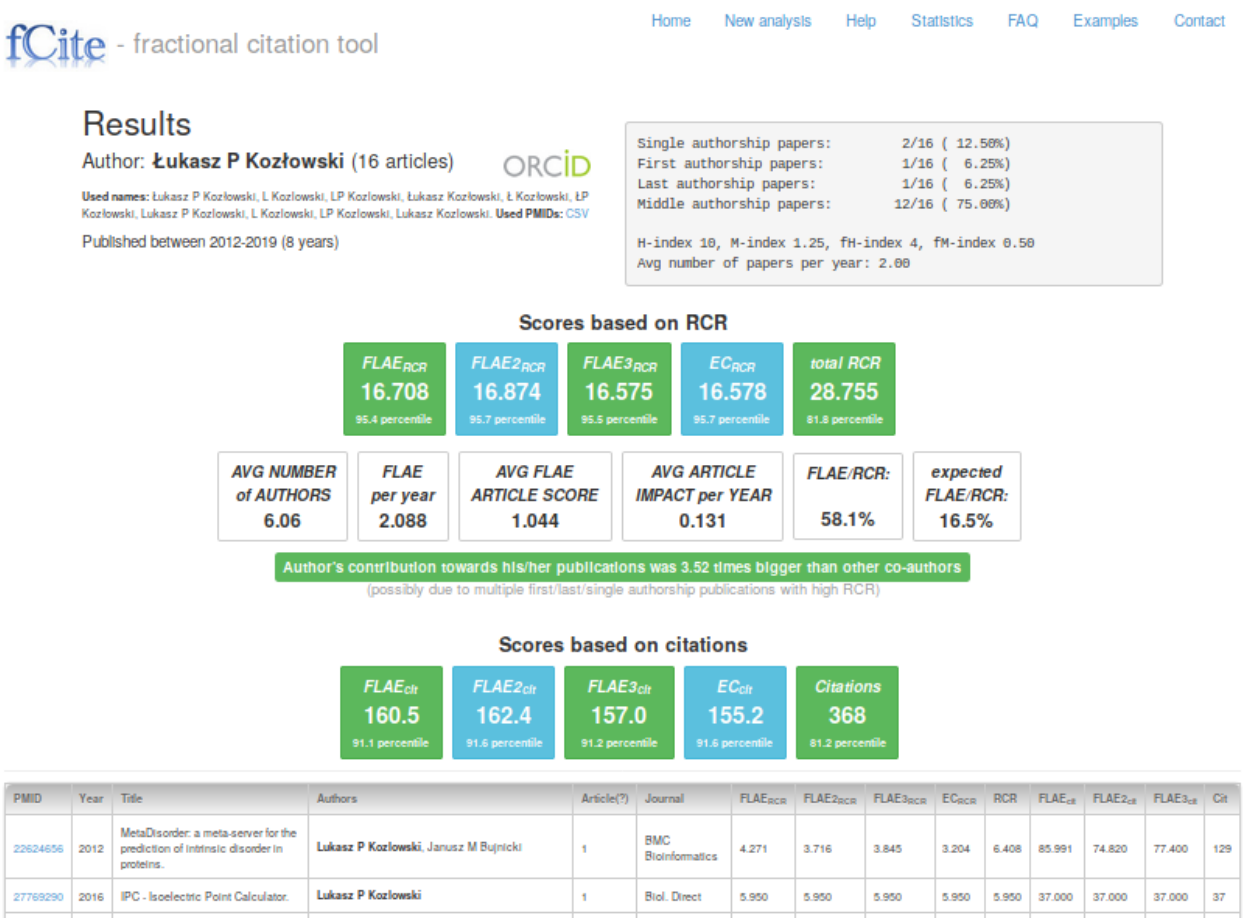
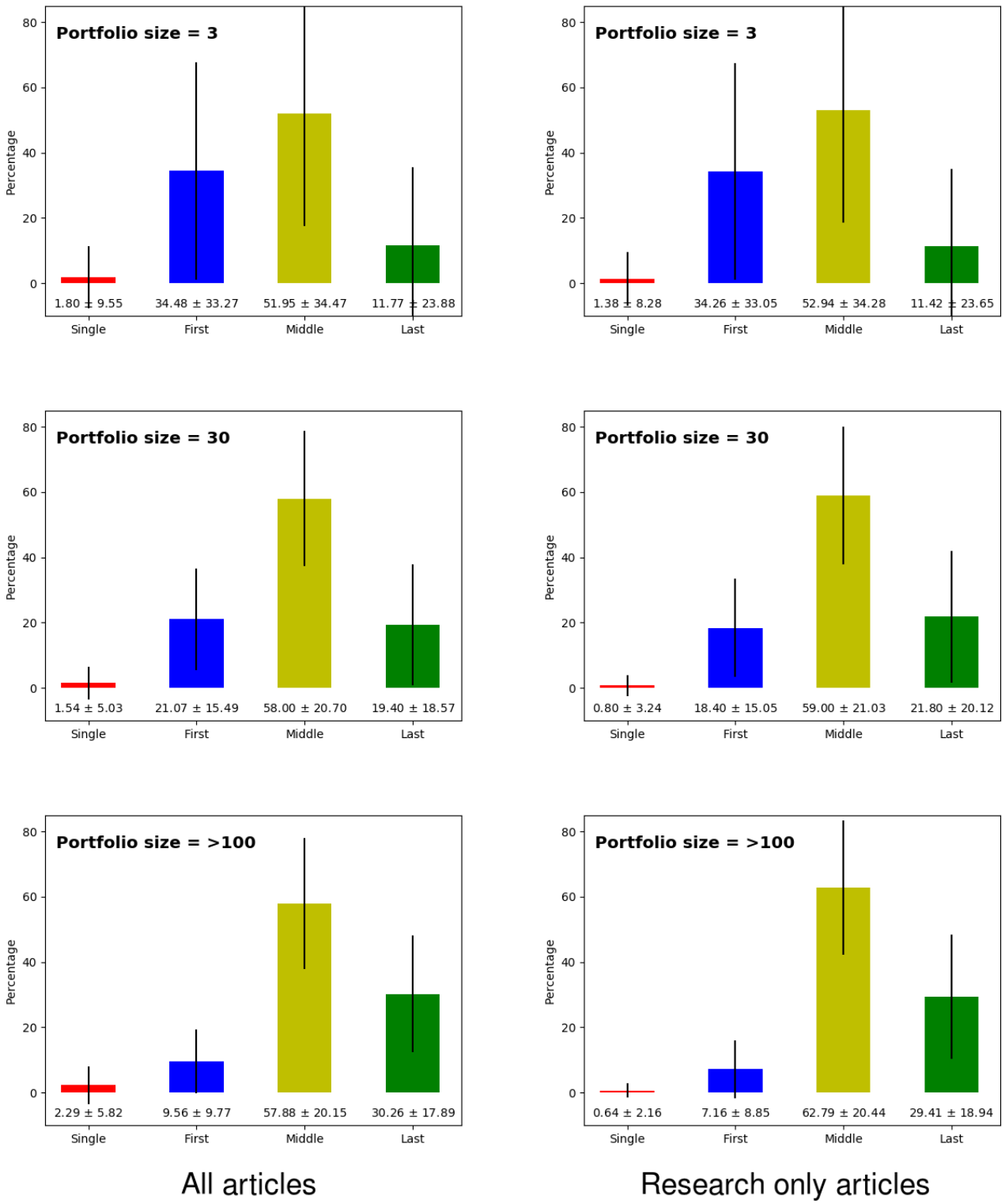


**Figure 5.** Percentage of single, first, middle and last authorship papers with respect to the size of the portfolio. For the numerical data, see **Supplementary Tables 8-9**. Animated version of the the figure is available at: http://fcite.org/stats.html#bar_plot



| | RCR | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | FLAE | | FLAE2 | | FLAE3 | | EC | | Total RCR | |
| | All | Research | All | Research | All | Research | All | Research | All | Research |
| 25.0 % | 0.241 | 0.222 | 0.265 | 0.247 | 0.259 | 0.241 | 0.253 | 0.236 | 1.470 | 1.398 |
| 50.0 % | 0.993 | 0.893 | 0.958 | 0.859 | 0.965 | 0.867 | 0.909 | 0.811 | 5.363 | 4.950 |
| 60.0 % | 1.628 | 1.436 | 1.541 | 1.353 | 1.560 | 1.374 | 1.462 | 1.277 | 8.809 | 8.035 |
| 70.0 % | 2.679 | 2.320 | 2.519 | 2.248 | 2.556 | 2.205 | 2.398 | 2.048 | 14.683 | 13.194 |
| 75.0 % | 3.475 | 2.984 | 3.263 | 2.776 | 3.315 | 2.825 | 3.118 | 2.634 | 19.203 | 17.145 |
| 80.0 % | 4.598 | 3.895 | 4.320 | 3.632 | 4.382 | 3.715 | 4.140 | 3.468 | 25.707 | 22.808 |
| 85.0 % | 6.318 | 5.247 | 5.922 | 4.892 | 6.009 | 4.982 | 5.713 | 4.694 | 35.766 | 31.348 |
| 90.0 % | 9.238 | 7.535 | 8.724 | 7.056 | 9.628 | 7.177 | 8.452 | 6.811 | 53.620 | 46.317 |
| 92.5 % | 11.733 | 9.417 | 11.126 | 8.868 | 11.297 | 9.041 | 10.827 | 8.625 | 69.633 | 59.395 |
| 95.0 % | 15.923 | 12.570 | 15.207 | 11.857 | 15.475 | 12.092 | 14.964 | 11.614 | 97.613 | 81.829 |
| 96.0 % | 18.662 | 14.551 | 17.894 | 13.810 | 18.163 | 14.101 | 17.628 | 16.149 | 116.382 | 96.418 |
| 97.0 % | 22.703 | 17.411 | 21.716 | 16.642 | 22.175 | 16.959 | 21.549 | 18.530 | 142.459 | 117.749 |
| 98.0 % | 29.433 | 22.257 | 28.435 | 21.334 | 28.951 | 21.776 | 28.227 | 21.270 | 190.610 | 150.524 |
| 99.0 % | 44.580 | 33.573 | 43.294 | 32.277 | 44.129 | 33.007 | 43.448 | 32.275 | 303.261 | 232.081 |
| 99.5 % | 68.146 | 51.340 | 66.220 | 49.257 | 67.798 | 50.669 | 66.144 | 49.132 | 474.222 | 349.833 |
| 99.6 % | 79.580 | 60.028 | 77.070 | 57.461 | 78.620 | 59.118 | 77.186 | 57.563 | 556.720 | 396.323 |
| 99.7 % | 99.570 | 77.268 | 95.486 | 75.146 | 97.869 | 76.975 | 96.026 | 75.933 | 704.544 | 459.216 |
| 99.8 % | 150.133 | 177.843 | 144.683 | 169.492 | 149.153 | 175.890 | 145.143 | 169.598 | 1087.509 | 575.450 |
| 99.9 % | 1359.852 | 1208.689 | 953.556 | 823.147 | 1120.820 | 995.152 | 798.151 | 639.006 | 5202.912 | 3998.950 |

| | Citations | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | All | | | | | Research | | | | |
| | FLAE | FLAE2 | FLAE3 | EC | Total | FLAE | FLAE2 | FLAE3 | EC | Total |
| 25.0 % | 1.3 | 1.4 | 1.3 | 1.3 | 7.0 | 1.2 | 1.3 | 1.3 | 1.2 | 7.0 |
| 50.0 % | 7.5 | 7.5 | 7.5 | 7.2 | 42.7 | 7.0 | 7.0 | 6.9 | 6.6 | 40.4 |
| 60.0 % | 14.7 | 14.5 | 14.5 | 13.9 | 83.4 | 13.4 | 13.1 | 13.2 | 12.5 | 77.7 |
| 70.0 % | 29.0 | 28.0 | 28.2 | 26.8 | 163.7 | 25.8 | 24.9 | 25.0 | 23.7 | 150.4 |
| 75.0 % | 41.2 | 39.5 | 39.8 | 37.8 | 232.7 | 36.4 | 34.7 | 35.0 | 33.0 | 211.8 |
| 80.0 % | 59.5 | 56.8 | 57.2 | 54.3 | 336.6 | 51.9 | 49.2 | 50.0 | 46.9 | 304.4 |
| 85.0 % | 88.9 | 84.5 | 85.4 | 81.2 | 506.0 | 76.1 | 76.0 | 72.8 | 68.7 | 452.2 |
| 90.0 % | 158.0 | 135.3 | 137.0 | 130.6 | 816.8 | 119.2 | 112.3 | 113.7 | 107.8 | 717.9 |
| 92.5 % | 191.6 | 181.0 | 183.8 | 175.0 | 1105.1 | 157.0 | 147.4 | 150.1 | 142.0 | 961.8 |
| 95.0 % | 273.9 | 259.6 | 263.6 | 252.8 | 1613.5 | 221.1 | 207.3 | 211.1 | 201.1 | 1378.5 |
| 96.0 % | 326.3 | 310.8 | 314.8 | 303.8 | 1965.3 | 261.2 | 245.2 | 250.2 | 238.4 | 1655.7 |
| 97.0 % | 406.8 | 388.3 | 394.7 | 380.4 | 2487.6 | 317.7 | 300.0 | 305.8 | 294.5 | 2079.5 |
| 98.0 % | 542.3 | 517.8 | 527.1 | 511.9 | 3401.0 | 414.5 | 394.2 | 402.3 | 386.6 | 2786.9 |
| 99.0 % | 850.0 | 819.8 | 833.5 | 808.0 | 5622.9 | 629.3 | 599.5 | 612.8 | 597.7 | 4472.3 |
| 99.5 % | 1301.2 | 1252.8 | 1278.0 | 1254.5 | 9546.5 | 956.2 | 902.7 | 927.3 | 902.8 | 7013.1 |
| 99.6 % | 1488.8 | 1442.5 | 1478.3 | 1440.8 | 11446.9 | 1098.8 | 1053.0 | 1081.6 | 1053.0 | 8093.1 |
| 99.7 % | 1798.4 | 1752.0 | 1780.5 | 1746.7 | 14926.8 | 1335.5 | 1275.7 | 1316.7 | 1276.2 | 9739.7 |
| 99.8 % | 2398.1 | 2329.2 | 2390.5 | 2337.4 | 25862.2 | 1820.2 | 1757.4 | 1787.8 | 1779.6 | 12812.4 |
| 99.9 % | 24012.0 | 20191.8 | 21141.2 | 20336.2 | 88650.7 | 21915.9 | 15234.5 | 18257.1 | 13323.2 | 85374.5 |

**TABLE 1.** Percentile scores based on ORCID profiles in 2018 for the key metrics used in *fCite* with respect to RCR and citations (a selection of the thresholds is presented; all data were bootstrapped 1000 times; and for complete list with the supporting values, see the files at http://www.fcite.org/percentiles_2018/). parts For the ratio and spread between fractional metrics (e.g., FLEA) and total RCR (Total Citations), see **Supplementary Figures 1-2**.

# Supplementary Materials for

## *fCite*: a fractional citation tool to quantify an individual's scientific research output

Lukasz P. Kozlowski
Correspondence to: lukasz.kozlowski.lpk@gmail.com

**This PDF file includes:**

Supplementary Text
Captions for Supplementary Data 1-3
Supplementary Figures 1-4
Supplementary Tables 1-10

**Supplementary Text**

Motivation

The number of the authors steadily increase over last 25 years (**Supplementary Table 5**). Moreover, when we divide the publications into research items (describing the original works) and non-research (e.g., the reviews, the editorials, etc.) we clearly can see that on average research publication require more authors. Given the fact that currently the publications with dozens or even hundreds of authors are common, it is desirable to modify the bibliometric metrics (e.g., citations, H-index) to seize the number of the authors for individual paper.

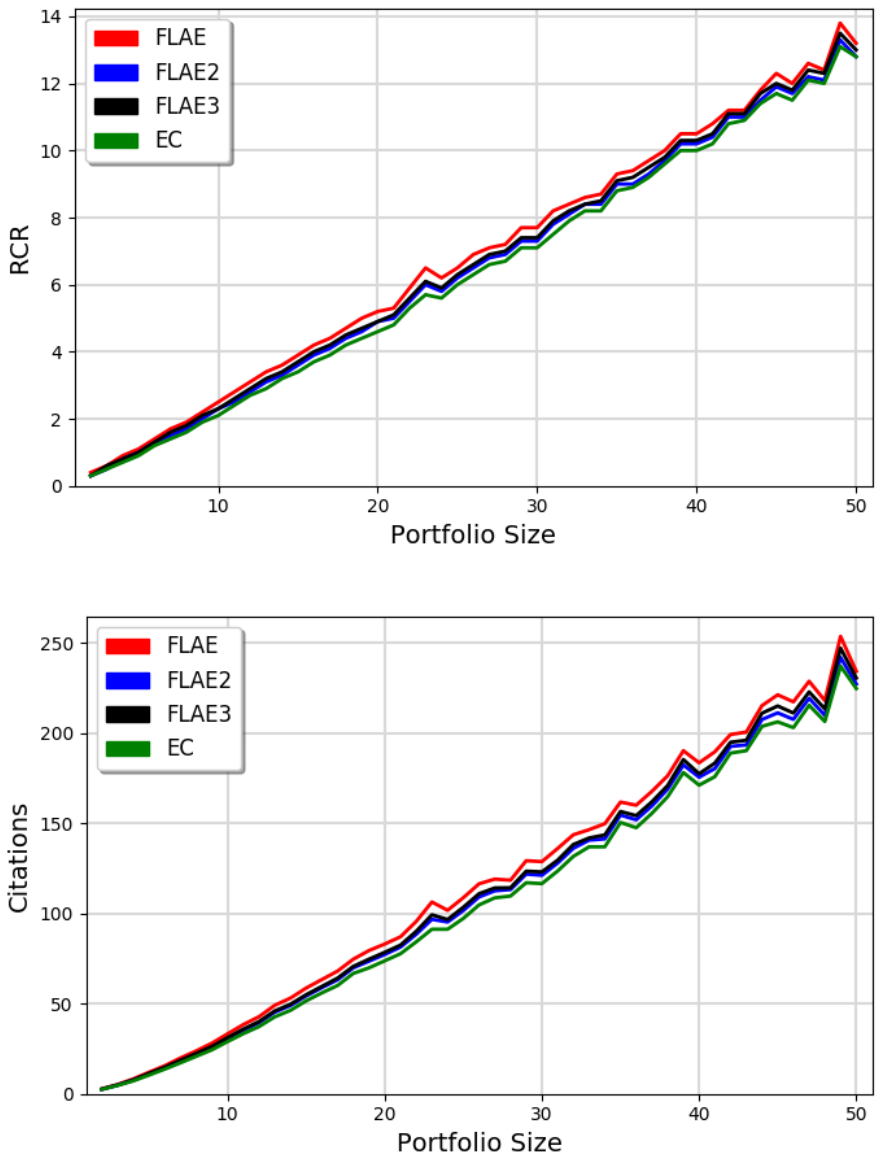
The patterns of authorship versus portfolio size

The ratio between fractional and total metric declines as the size of portfolio increase (**Supplementary Fig. 1**). Regardless of the main metric used (RCR or citation) the trend is stable and the data show that small portfolios have more first author publications. When the portfolio size increases more last author items appear. Depending the fractional model used, the small portfolios have 20-25% of contribution of the author falling below 18% for bigger portfolios. Here, it is interesting to point that regardless of fractional model (FLAE vs FLAE2 vs FLAE3 vs EC) the portfolios with around 40-60 items score virtually identical (which means that for mature scientist with multiple publications is not important which fractional model will be used). One should also take into account the spread of the scores which is huge for small size portfolios and decreases over the number of items, but it is always very significant and comprise at least roughly 20% (**Supplementary Fig. 2**).

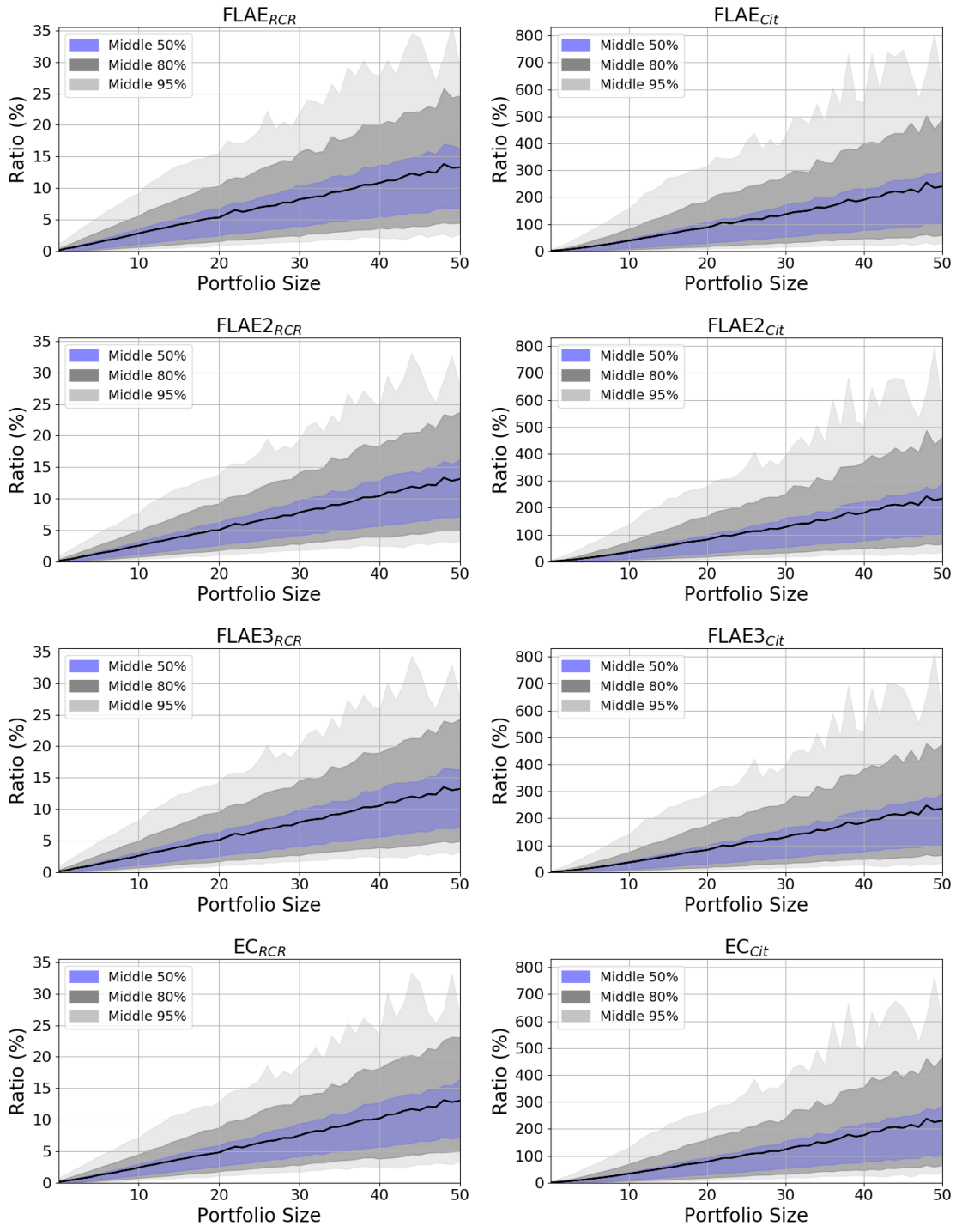
**Supplementary Data 1.** PUBMED data set containing publications with PMID numbers from 7 million to 19 million (json format). http://www.fcite.org/icite_1218_7M-19M.tar.gz or http://dx.doi.org/10.18150/repod.3945420

**Supplementary Data 2.** PUBMED data set containing publications with PMID numbers from 20 million to 32 million (json format). http://www.fcite.org/icite_1218_20M-32M.tar.gz or http://dx.doi.org/10.18150/repod.2195699

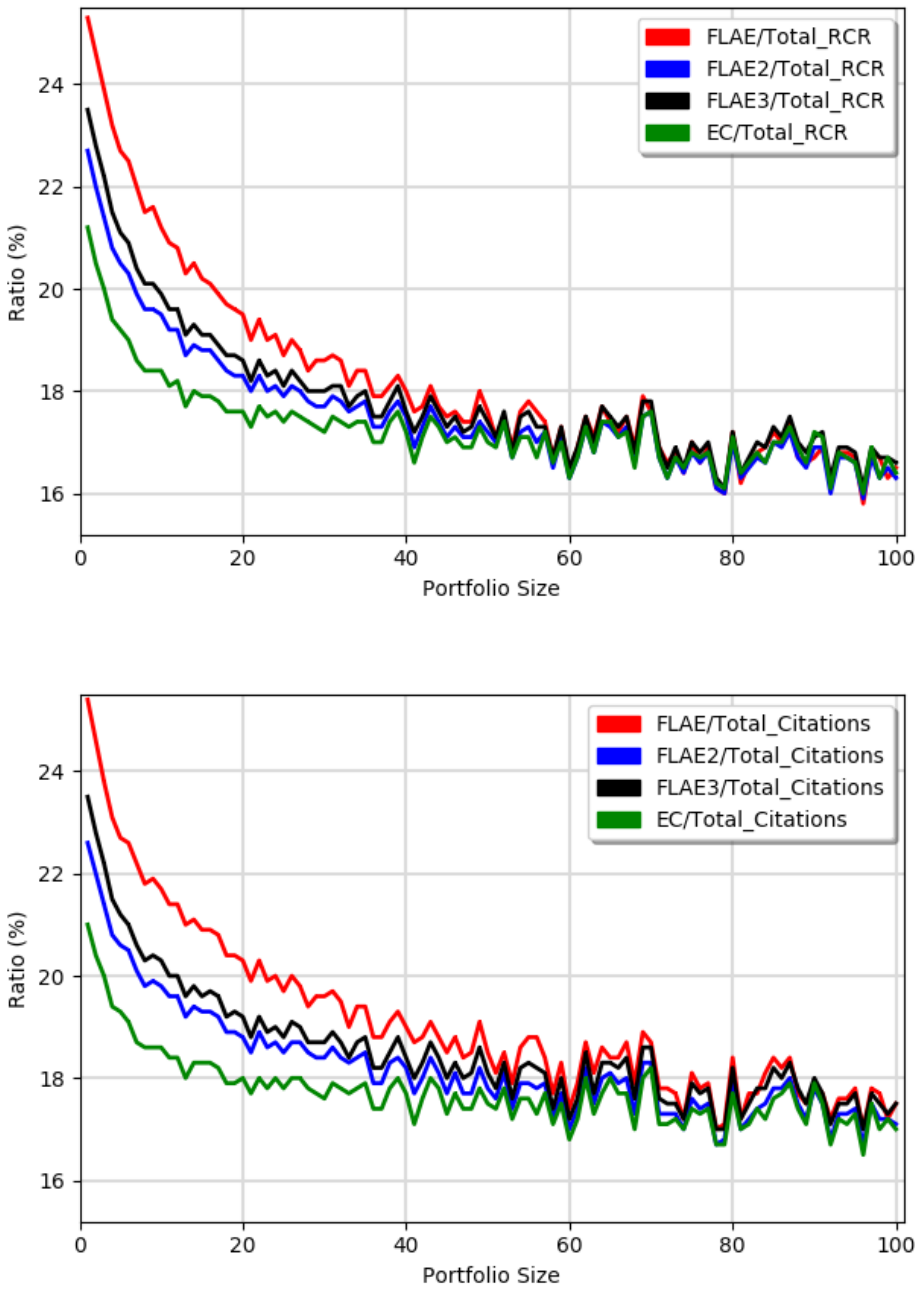
**Supplementary Data 3.** ORCID Public Data File (xml format). https://figshare.com/articles/ORCID_Public_Data_File_2018/7234028



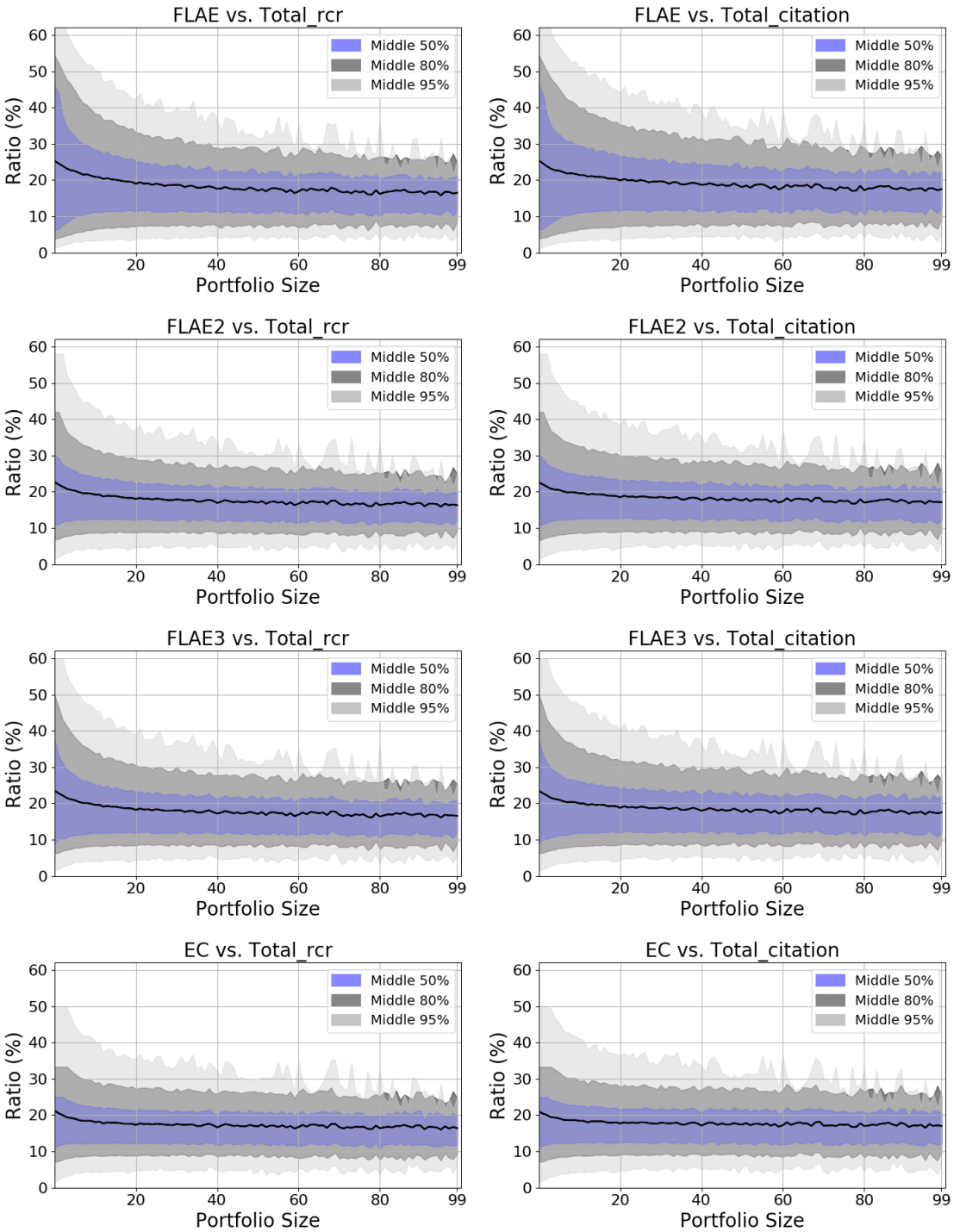

**Supplementary Figure 1.** Fractional metrics ($FLAE_{RCR}$, $FLAE_{Cit}$, $FLAE2_{RCR}$, $FLAE2_{Cit}$, $FLAE3_{RCR}$, $FLAE3_{Cit}$, $EC_{RCR}$, $EC_{Cit}$) in respect to the portfolio size (only the 394,189 ORCID portfolios with 2-50 items are presented).



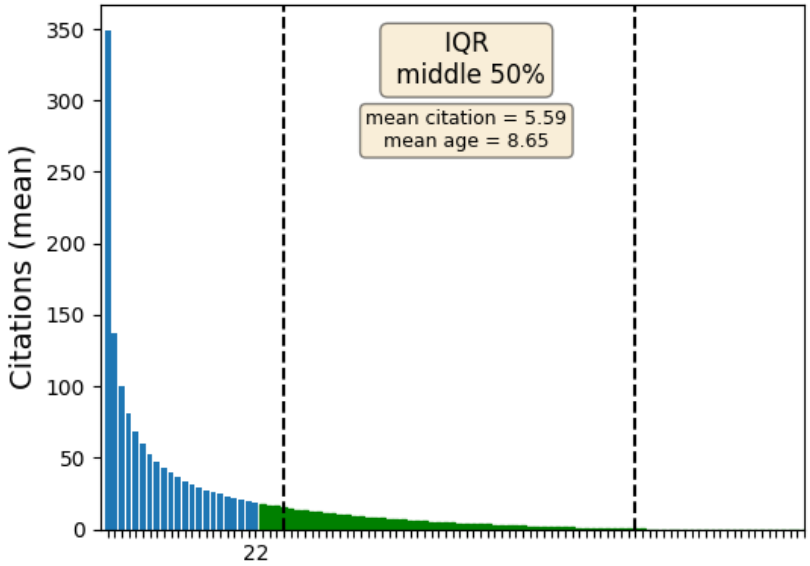

**Supplementary Figure 2.** The spread of the fractional metrics ($FLAE_{RCR}$, $FLAE_{Cit}$, $FLAE2_{RCR}$, $FLAE2_{Cit}$, $FLAE3_{RCR}$, $FLAE3_{Cit}$, $EC_{RCR}$, $EC_{Cit}$) in respect to the portfolio size. For the animated plots see fcite.org/stats.html



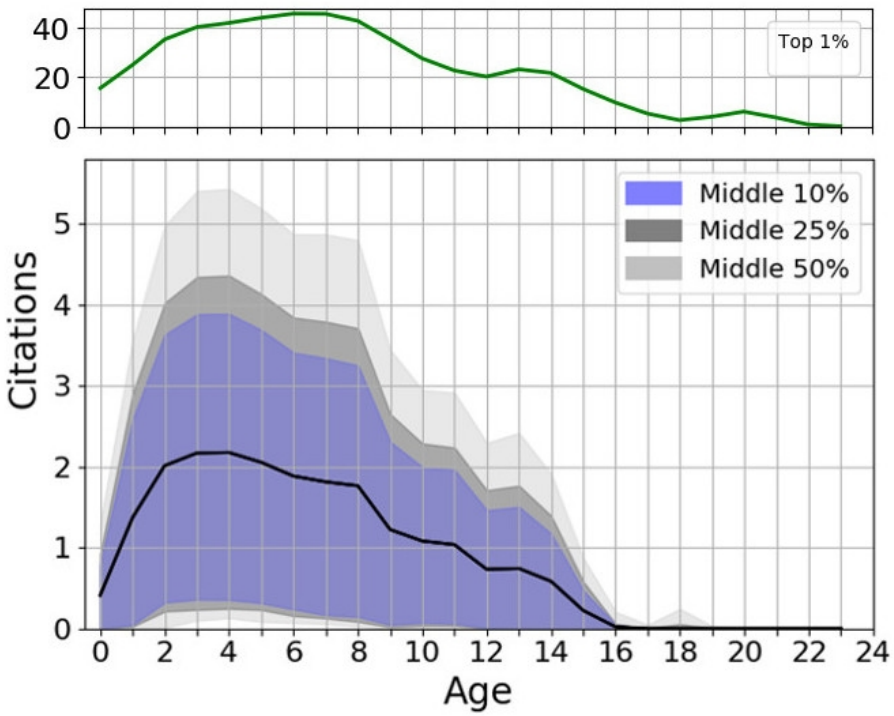

**Supplementary Figure 3.** The ratio between the fractional metrics (FLAE, FLAE2, FLAE3, EC) vs. total metrics (RCR or Citations) in respect to the portfolio size.




**Supplementary Figure 4.** The spread of the ratio between the fractional metrics (FLAE, FLAE2, FLAE3, EC) vs. total metrics (RCR or Citations) in respect to the portfolio size. For the animated plots see htt://www.fcite.org/stats.html




**Supplementary Figure 5.** Non-Gaussian distribution of the citations received by the publications (Lotka's low). Based on 17,787,017 PUBMED publications that received 288,085,850 citations (research and non-research items) within 1995-2018. Blue bars correspond to 80% of the citations (22$^{nd}$ percent of the publications obtained 80% of all citations, in the agreement with Pareto principle). The average publication (within 50% middle range) is expected to obtain only 5-6 citations during 8 years period.




**Supplementary Figure 6.** Time dependency of the citations with respect to the age of the article. The citation distribution is highly positively skewed, most of the citations are received within a 2-5 year period followed with long right tail thwarting after 15 years (bottom panel). The similar shape of the distribution can be observed even for the top 1% of best-cited publications, but some differences can be observed. First of all, the number of citations is ~20 times bigger than on average (note different y scales on both panels). Next, the right tail is much longer and the distribution is less skewed.

**SUPPLEMENTARY TABLE 1.** The confusion matrix for up to ten authors for FLAE model. A model with strong emphasis of the first and the last author. For the more details see: http://www.fcite.org/FLAE.txt

| No. | Weights for individual authors |
|---|---|
| 1 | 1.0000 |
| 2 | 0.6667, 0.3333 |
| 3 | 0.5455, 0.1818, 0.2727 |
| 4 | 0.5000, 0.1250, 0.1250, 0.2500 |
| 5 | 0.4762, 0.0952, 0.0952, 0.0952, 0.2381 |
| 6 | 0.4615, 0.0769, 0.0769, 0.0769, 0.0769, 0.2308 |
| 7 | 0.4516, 0.0645, 0.0645, 0.0645, 0.0645, 0.0645, 0.2258 |
| 8 | 0.4444, 0.0556, 0.0556, 0.0556, 0.0556, 0.0556, 0.0556, 0.2222 |
| 9 | 0.4390, 0.0488, 0.0488, 0.0488, 0.0488, 0.0488, 0.0488, 0.0488, 0.2195 |
| 10 | 0.4348, 0.0435, 0.0435, 0.0435, 0.0435, 0.0435, 0.0435, 0.0435, 0.0435, 0.2174 |

**SUPPLEMENTARY TABLE 2.** The confusion matrix for up to ten authors for FLAE2 model (based on (5)). A model with moderate emphasis of the first and the last author. For the more details see: http://www.fcite.org/FLAE2.txt

| No. | Weights for individual authors |
|---|---|
| 1 | 1.0000 |
| 2 | 0.5800, 0.4200 |
| 3 | 0.4200, 0.2700, 0.3100 |
| 4 | 0.3500, 0.2200, 0.1800, 0.2500 |
| 5 | 0.3000, 0.1950, 0.1450, 0.1450, 0.2150 |
| 6 | 0.2600, 0.1700, 0.1300, 0.1200, 0.1200, 0.2000 |
| 7 | 0.2300, 0.1600, 0.1200, 0.1100, 0.1000, 0.1000, 0.1800 |
| 8 | 0.2100, 0.1500, 0.1100, 0.0950, 0.0900, 0.0900, 0.0850, 0.1700 |
| 9 | 0.1950, 0.1400, 0.1000, 0.0850, 0.0820, 0.0810, 0.0790, 0.0780, 0.1600 |
| 10 | 0.1800, 0.1300, 0.0900, 0.0820, 0.0780, 0.0750, 0.0730, 0.0720, 0.0700, 0.150 |

**SUPPLEMENTARY TABLE 3.** The confusion matrix for up to ten authors for FLAE3 model. For the more details see: http://www.fcite.org/FLAE3.txt

| No. | Weights for individual authors |
|---|---|
| 1 | 1.0 |
| 2 | 0.6000, 0.4000 |
| 3 | 0.5000, 0.1667, 0.3333 |
| 4 | 0.4286, 0.1429, 0.1429, 0.2857 |
| 5 | 0.3750, 0.1250, 0.1250, 0.1250, 0.2500 |
| 6 | 0.3333, 0.1111, 0.1111, 0.1111, 0.1111, 0.2222 |
| 7 | 0.3000, 0.1000, 0.1000, 0.1000, 0.1000, 0.1000, 0.2000 |
| 8 | 0.2727, 0.0909, 0.0909, 0.0909, 0.0909, 0.0909, 0.0909, 0.1818 |
| 9 | 0.2500, 0.0833, 0.0833, 0.0833, 0.0833, 0.0833, 0.0833, 0.0833, 0.1667 |
| 10 | 0.2308, 0.0769, 0.0769, 0.0769, 0.0769, 0.0769, 0.0769, 0.0769, 0.0769, 0.1538 |

**SUPPLEMENTARY TABLE 4.** The confusion matrix for up to ten authors for EC model (equal contribution of all authors). For the more details see: http://www.fcite.org/EC.txt

| No. | Weights for individual authors |
|---|---|
| 1 | 1.0000 |
| 2 | 0.5000, 0.5000 |
| 3 | 0.3333, 0.3333, 0.3333 |
| 4 | 0.2500, 0.2500, 0.2500, 0.2500 |
| 5 | 0.2000, 0.2000, 0.2000, 0.2000, 0.2000 |
| 6 | 0.1667, 0.1667, 0.1667, 0.1667, 0.1667, 0.1667 |
| 7 | 0.1429, 0.1429, 0.1429, 0.1429, 0.1429, 0.1429, 0.1429 |
| 8 | 0.1250, 0.1250, 0.1250, 0.1250, 0.1250, 0.1250, 0.1250, 0.1250 |
| 9 | 0.1111, 0.1111, 0.1111, 0.1111, 0.1111, 0.1111, 0.1111, 0.1111, 0.1111 |
| 10 | 0.1000, 0.1000, 0.1000, 0.1000, 0.1000, 0.1000, 0.1000, 0.1000, 0.1000, 0.1000 |

**SUPPLEMENTARY TABLE 5.** The correlations between fractional models and total scores. The lower triangular portion of the matrices (green) correspond to 394,189 ORCID portfolios with 2-50 items and the upper triangular portion of the matrices correspond to all ORCID portfolios with at least single item (600,755 portfolios).

| | RCR | | | | | | Citations | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **FLAE** | **FLAE2** | **FLAE3** | **EC** | **Total** | | **FLAE** | **FLAE2** | **FLAE3** | **EC** | **Total** |
| **FLAE** | | 0.9852 | 0.9929 | 0.9652 | 0.7592 | **FLAE** | | 0.9874 | 0.9937 | 0.9708 | 0.7883 |
| **FLAE2** | 0.9700 | | 0.9977 | 0.9951 | 0.7802 | **FLAE2** | 0.9741 | | 0.9980 | 0.9959 | 0.8051 |
| **FLAE3** | 0.9884 | 0.9937 | | 0.9876 | 0.7724 | **FLAE3** | 0.9896 | 0.9945 | | 0.9898 | 0.7985 |
| **EC** | 0.9237 | 0.9879 | 0.9675 | | 0.7837 | **EC** | 0.9363 | 0.9901 | 0.9730 | | 0.8073 |
| **Total** | 0.5762 | 0.6166 | 0.6012 | 0.6264 | | **Total** | 0.6466 | 0.6850 | 0.6703 | 0.6937 | |

**SUPPLEMENTARY TABLE 6.** The average number of the authors (PUBMED 17,787,016 publications in 1995-2018).

| Year | Research | Non-research | Research Avg | Not-research Avg | Total Avg |
|---|---|---|---|---|---|
| 1995 | 363007 | 85576 | 3.8849 | 2.0349 | 3.5320 |
| 1996 | 370198 | 87896 | 4.0317 | 2.0976 | 3.6606 |
| 1997 | 363577 | 92127 | 4.1102 | 2.0864 | 3.7011 |
| 1998 | 378481 | 95424 | 4.1766 | 2.0980 | 3.7581 |
| 1999 | 390468 | 102615 | 4.2272 | 2.0834 | 3.7811 |
| 2000 | 421771 | 107840 | 4.2740 | 2.1196 | 3.8353 |
| 2001 | 429734 | 113286 | 4.4378 | 2.1117 | 3.9525 |
| 2002 | 439721 | 118445 | 4.5258 | 2.1878 | 4.0297 |
| 2003 | 457320 | 126100 | 4.6535 | 2.2453 | 4.1330 |
| 2004 | 487038 | 132345 | 4.7827 | 2.3123 | 4.2549 |
| 2005 | 521093 | 134582 | 4.8916 | 2.3505 | 4.3700 |
| 2006 | 545681 | 138391 | 4.9606 | 2.4008 | 4.4427 |
| 2007 | 569964 | 139565 | 5.0582 | 2.4892 | 4.5529 |
| 2008 | 606463 | 143759 | 5.1299 | 2.5105 | 4.6280 |
| 2009 | 638608 | 144718 | 5.2028 | 2.5767 | 4.7176 |
| 2010 | 672589 | 149643 | 5.3194 | 2.6699 | 4.8372 |
| 2011 | 718523 | 156326 | 5.4893 | 2.7427 | 4.9985 |
| 2012 | 775637 | 163479 | 5.6627 | 2.8384 | 5.1710 |
| 2013 | 811789 | 181971 | 5.7698 | 2.9523 | 5.2539 |
| 2014 | 845090 | 195823 | 5.8954 | 3.0620 | 5.3624 |
| 2015 | 878182 | 209943 | 6.0838 | 3.2217 | 5.5316 |
| 2016 | 896432 | 209084 | 6.2214 | 3.4351 | 5.6945 |
| 2017 | 935012 | 189792 | 6.2433 | 3.6815 | 5.8111 |
| 2018 | 927277 | 123246 | 6.1820 | 3.8971 | 5.9139 |
| **Total** | 14444982 | 3342034 | 5.2915 | 2.6970 | 4.8040 |

**SUPPLEMENTARY TABLE 7.** The number of the authors over the years with the respect to the research and non-research items (years 1995-2018).

| Author number | All | Not research | Research | Non research fr | Fr from all |
|---|---|---|---|---|---|
| 1 | 2524210 | 1297946 | 1226264 | 51.4 | 14.2 |
| 2 | 2456770 | 800840 | 1655930 | 32.6 | 13.8 |
| 3 | 2562334 | 477521 | 2084813 | 18.6 | 14.4 |
| 4 | 2429355 | 280806 | 2148549 | 11.6 | 13.7 |
| 5 | 2089498 | 171352 | 1918146 | 8.2 | 11.7 |
| 6 | 1719482 | 111165 | 1608317 | 6.5 | 9.7 |
| 7 | 1203690 | 65535 | 1138155 | 5.4 | 6.8 |
| 8 | 858575 | 41505 | 817070 | 4.8 | 4.8 |
| 9 | 580374 | 25882 | 554492 | 4.5 | 3.3 |
| 10 | 425535 | 18655 | 406880 | 4.4 | 2.4 |
| 11 | 264146 | 11718 | 252428 | 4.4 | 1.5 |
| 12 | 186533 | 8447 | 178086 | 4.5 | 1.0 |
| 13 | 122871 | 6002 | 116869 | 4.9 | 0.7 |
| 14 | 8663 | 4482 | 82211 | 5.2 | 0.5 |
| 15 | 6279 | 3421 | 59373 | 5.4 | 0.4 |
| >15 | 212771 | 16699 | 196072 | 7.8 | 1.2 |

**SUPPLEMENTARY TABLE 8.** Percentage of single-, first-, middle- and last-authorship papers in respect to the size of portfolio for all articles (both research and non-research items). As the number of available portfolios decreases as the size of portfolio increases the data has been bootstrapped 10,000 times (especially crucial for the portfolios with >50 items, where there is less than 1,000 of portfolios passing the threshold).

| Portfolio size | Single | First | Middle | Last | No. portfolios |
|---|---|---|---|---|---|
| 3 | 1.805 ± 9.548 | 34.475 ± 33.269 | 51.954 ± 34.474 | 11.766 ± 23.878 | 45339 |
| 4 | 1.600 ± 8.346 | 33.820 ± 30.054 | 53.406 ± 31.352 | 11.174 ± 21.729 | 33626 |
| 5 | 1.555 ± 7.716 | 33.299 ± 27.893 | 54.422 ± 29.307 | 10.723 ± 20.373 | 26515 |
| 6 | 1.544 ± 7.499 | 32.813 ± 26.068 | 55.360 ± 27.398 | 10.283 ± 19.009 | 21897 |
| 7 | 1.383 ± 6.594 | 32.119 ± 24.739 | 56.157 ± 26.255 | 10.341 ± 18.240 | 18389 |
| 8 | 1.383 ± 6.394 | 31.464 ± 23.650 | 56.643 ± 25.286 | 10.510 ± 17.947 | 16033 |
| 9 | 1.458 ± 6.596 | 31.213 ± 22.931 | 56.310 ± 24.484 | 11.019 ± 17.963 | 13892 |
| 10 | 1.496 ± 6.582 | 30.254 ± 22.022 | 56.988 ± 24.254 | 11.262 ± 17.676 | 12545 |
| 11 | 1.348 ± 5.868 | 29.824 ± 21.263 | 57.296 ± 23.429 | 11.533 ± 17.464 | 10674 |
| 12 | 1.494 ± 6.325 | 29.308 ± 20.851 | 57.367 ± 23.050 | 11.831 ± 17.220 | 9653 |
| 13 | 1.319 ± 5.260 | 28.420 ± 20.173 | 58.037 ± 22.475 | 12.223 ± 17.384 | 8761 |
| 14 | 1.497 ± 5.700 | 28.102 ± 19.678 | 57.624 ± 22.541 | 12.777 ± 17.546 | 7963 |
| 15 | 1.518 ± 5.847 | 27.034 ± 19.400 | 57.899 ± 22.441 | 13.549 ± 17.865 | 7303 |
| 16 | 1.419 ± 5.177 | 26.517 ± 18.750 | 57.879 ± 21.824 | 14.186 ± 18.134 | 6690 |
| 17 | 1.515 ± 6.038 | 26.133 ± 18.512 | 58.186 ± 21.764 | 14.166 ± 17.848 | 6093 |
| 18 | 1.548 ± 5.985 | 25.533 ± 17.812 | 58.381 ± 21.498 | 14.538 ± 17.850 | 5589 |
| 19 | 1.413 ± 5.369 | 25.220 ± 17.959 | 57.909 ± 21.742 | 15.459 ± 18.196 | 5231 |
| 20 | 1.499 ± 5.189 | 24.730 ± 17.712 | 57.797 ± 21.450 | 15.974 ± 18.465 | 4914 |
| 21 | 1.495 ± 5.488 | 23.655 ± 17.131 | 58.381 ± 21.386 | 16.470 ± 18.295 | 4513 |
| 22 | 1.618 ± 5.738 | 23.215 ± 16.769 | 57.678 ± 20.800 | 17.489 ± 18.605 | 4226 |
| 23 | 1.395 ± 4.673 | 22.913 ± 16.182 | 58.659 ± 20.672 | 17.033 ± 18.196 | 3835 |
| 24 | 1.509 ± 5.018 | 22.809 ± 16.454 | 58.410 ± 20.849 | 17.272 ± 18.156 | 3591 |
| 25 | 1.587 ± 5.574 | 21.824 ± 15.937 | 57.868 ± 20.929 | 18.722 ± 19.077 | 3515 |
| 26 | 1.683 ± 5.586 | 21.637 ± 15.999 | 57.910 ± 21.264 | 18.770 ± 18.792 | 3117 |
| 27 | 1.526 ± 4.826 | 21.493 ± 15.934 | 57.873 ± 21.028 | 19.109 ± 19.375 | 3009 |
| 28 | 1.753 ± 5.803 | 20.772 ± 15.170 | 58.115 ± 20.745 | 19.361 ± 18.690 | 2677 |
| 29 | 1.591 ± 5.079 | 21.462 ± 15.548 | 57.553 ± 20.608 | 19.394 ± 18.521 | 2635 |
| 30 | 1.536 ± 5.026 | 21.069 ± 15.493 | 57.997 ± 20.703 | 19.399 ± 18.568 | 2471 |
| 31 | 1.619 ± 5.047 | 20.566 ± 15.645 | 57.736 ± 20.884 | 20.079 ± 18.652 | 2309 |
| 32 | 1.429 ± 4.202 | 19.410 ± 14.861 | 58.121 ± 20.370 | 21.040 ± 18.852 | 2140 |
| 33 | 1.633 ± 5.016 | 19.326 ± 14.633 | 58.249 ± 20.202 | 20.792 ± 18.749 | 2137 |
| 34 | 1.694 ± 4.735 | 19.348 ± 14.385 | 57.530 ± 20.097 | 21.429 ± 18.645 | 1907 |
| 35 | 1.712 ± 5.219 | 19.479 ± 15.002 | 57.346 ± 20.390 | 21.463 ± 18.872 | 1810 |
| 36 | 1.586 ± 4.191 | 18.528 ± 13.804 | 57.546 ± 20.385 | 22.340 ± 18.958 | 1700 |
| 37 | 1.617 ± 4.168 | 18.454 ± 14.539 | 56.617 ± 20.446 | 23.312 ± 19.643 | 1729 |
| 38 | 1.646 ± 4.668 | 17.855 ± 13.562 | 56.644 ± 19.811 | 23.855 ± 19.271 | 1543 |
| 39 | 1.743 ± 4.525 | 17.784 ± 14.104 | 55.639 ± 20.514 | 24.834 ± 19.230 | 1489 |
| 40 | 1.640 ± 5.088 | 18.365 ± 14.100 | 57.916 ± 19.948 | 22.080 ± 18.400 | 1461 |
| 41 | 1.559 ± 4.733 | 17.680 ± 13.569 | 57.068 ± 20.432 | 23.693 ± 18.949 | 1305 |
| 42 | 1.660 ± 4.586 | 17.175 ± 13.710 | 57.428 ± 21.059 | 23.736 ± 19.635 | 1367 |
| 43 | 1.666 ± 3.887 | 17.124 ± 13.465 | 56.138 ± 20.564 | 25.072 ± 19.799 | 1186 |
| 44 | 1.939 ± 6.258 | 16.137 ± 12.661 | 57.008 ± 20.561 | 24.916 ± 19.268 | 1172 |
| 45 | 1.639 ± 3.895 | 16.645 ± 12.841 | 57.957 ± 19.972 | 23.759 ± 18.791 | 1137 |
| 46 | 1.503 ± 3.613 | 16.686 ± 12.620 | 57.234 ± 20.118 | 24.577 ± 19.090 | 1082 |
| 47 | 1.547 ± 4.454 | 16.028 ± 13.354 | 57.871 ± 20.025 | 24.553 ± 19.081 | 1071 |
| 48 | 1.608 ± 4.321 | 16.422 ± 13.348 | 56.813 ± 20.267 | 25.157 ± 19.618 | 995 |
| 49 | 1.961 ± 4.995 | 17.026 ± 14.054 | 54.989 ± 20.002 | 26.023 ± 19.448 | 872 |
| 50 | 1.960 ± 5.016 | 15.880 ± 13.182 | 57.866 ± 21.060 | 24.294 ± 18.893 | 844 |
| 51 | 2.191 ± 6.857 | 15.464 ± 12.934 | 57.460 ± 19.505 | 24.885 ± 18.349 | 833 |
| 52 | 2.015 ± 5.349 | 14.681 ± 12.171 | 57.543 ± 20.084 | 25.760 ± 18.946 | 790 |
| 53 | 1.815 ± 5.731 | 14.966 ± 12.815 | 57.212 ± 19.556 | 26.007 ± 18.938 | 773 |
| 54 | 1.971 ± 5.910 | 15.394 ± 13.198 | 56.118 ± 20.057 | 26.516 ± 19.112 | 714 |
| 55 | 1.963 ± 5.130 | 15.330 ± 12.747 | 55.771 ± 20.303 | 26.936 ± 19.267 | 702 |
| 56 | 1.928 ± 4.848 | 16.255 ± 14.003 | 54.931 ± 20.663 | 26.887 ± 19.520 | 689 |
| 57 | 1.888 ± 4.978 | 14.467 ± 12.267 | 56.794 ± 19.314 | 26.851 ± 18.204 | 665 |
| 58 | 1.715 ± 5.420 | 12.023 ± 11.994 | 63.210 ± 22.986 | 23.053 ± 20.018 | 743 |
| 59 | 2.002 ± 4.978 | 14.772 ± 12.344 | 56.207 ± 20.338 | 27.020 ± 19.037 | 625 |
| 60 | 2.016 ± 5.181 | 13.658 ± 11.762 | 56.921 ± 19.855 | 27.406 ± 19.277 | 596 |
| 61 | 1.997 ± 5.474 | 14.290 ± 12.254 | 56.668 ± 20.230 | 27.046 ± 19.160 | 564 |
| 62 | 2.181 ± 5.761 | 13.878 ± 12.306 | 55.356 ± 20.023 | 28.585 ± 19.948 | 593 |
| 63 | 2.035 ± 5.975 | 13.834 ± 12.589 | 57.733 ± 19.533 | 26.397 ± 18.374 | 513 |
| 64 | 1.942 ± 5.172 | 14.811 ± 12.808 | 55.491 ± 20.259 | 27.756 ± 19.157 | 484 |
| 65 | 2.125 ± 5.615 | 13.566 ± 11.884 | 54.029 ± 20.029 | 30.280 ± 19.507 | 477 |
| 66 | 1.819 ± 3.345 | 13.863 ± 12.170 | 56.334 ± 19.336 | 27.984 ± 18.813 | 480 |
| 67 | 1.987 ± 4.412 | 14.851 ± 11.950 | 54.995 ± 19.528 | 28.168 ± 19.029 | 464 |
| 68 | 1.614 ± 3.076 | 13.586 ± 11.572 | 55.001 ± 19.257 | 29.798 ± 18.689 | 441 |
| 69 | 2.627 ± 7.300 | 13.604 ± 11.599 | 54.233 ± 20.245 | 29.536 ± 19.593 | 443 |
| 70 | 2.094 ± 4.409 | 13.313 ± 10.981 | 54.384 ± 19.327 | 30.209 ± 18.306 | 446 |
| 71 | 2.130 ± 5.856 | 13.517 ± 11.445 | 56.087 ± 20.005 | 28.267 ± 18.747 | 420 |
| 72 | 1.970 ± 4.943 | 13.006 ± 12.200 | 56.742 ± 20.123 | 28.282 ± 18.629 | 378 |
| 73 | 1.964 ± 5.752 | 13.368 ± 12.086 | 56.824 ± 19.408 | 27.845 ± 19.094 | 386 |
| 74 | 1.913 ± 4.216 | 13.206 ± 11.864 | 57.473 ± 19.135 | 27.407 ± 18.432 | 364 |
| 75 | 2.716 ± 9.263 | 12.496 ± 11.310 | 56.160 ± 21.150 | 28.627 ± 18.943 | 389 |
| 76 | 1.756 ± 3.237 | 12.001 ± 9.543 | 56.824 ± 18.841 | 29.420 ± 18.866 | 342 |
| 77 | 1.584 ± 3.346 | 13.373 ± 12.150 | 56.318 ± 19.939 | 28.725 ± 18.627 | 321 |
| 78 | 1.775 ± 4.549 | 10.771 ± 10.555 | 57.105 ± 20.530 | 30.350 ± 19.457 | 332 |
| 79 | 1.963 ± 3.918 | 12.348 ± 10.690 | 59.079 ± 19.993 | 26.609 ± 17.985 | 298 |
| 80 | 2.255 ± 6.448 | 12.759 ± 10.706 | 55.678 ± 19.228 | 29.308 ± 18.376 | 303 |
| 81 | 2.154 ± 4.787 | 11.919 ± 12.474 | 56.828 ± 20.363 | 29.099 ± 18.865 | 289 |
| 82 | 2.307 ± 5.251 | 12.537 ± 10.945 | 55.631 ± 19.251 | 29.525 ± 17.320 | 275 |
| 83 | 2.084 ± 3.601 | 11.856 ± 10.331 | 54.594 ± 19.476 | 31.466 ± 19.081 | 279 |
| 84 | 2.413 ± 5.634 | 13.262 ± 12.046 | 54.214 ± 19.589 | 30.111 ± 17.598 | 250 |
| 85 | 2.906 ± 6.011 | 12.006 ± 11.142 | 53.307 ± 19.023 | 31.781 ± 18.434 | 277 |
| 86 | 1.935 ± 6.884 | 12.048 ± 10.737 | 56.640 ± 18.970 | 29.376 ± 18.088 | 243 |
| 87 | 2.948 ± 8.689 | 12.396 ± 11.480 | 53.846 ± 19.963 | 30.810 ± 18.548 | 267 |
| 88 | 1.876 ± 3.957 | 11.642 ± 11.661 | 55.477 ± 20.183 | 31.005 ± 18.398 | 241 |
| 89 | 2.014 ± 3.353 | 12.898 ± 11.243 | 53.572 ± 18.834 | 31.515 ± 18.705 | 209 |
| 90 | 2.068 ± 5.421 | 11.075 ± 9.559 | 56.342 ± 17.139 | 30.515 ± 17.036 | 204 |
| 91 | 2.726 ± 7.159 | 10.376 ± 9.813 | 54.145 ± 19.719 | 32.753 ± 18.997 | 226 |
| 92 | 1.840 ± 3.979 | 11.475 ± 10.994 | 56.182 ± 19.041 | 30.503 ± 18.615 | 212 |
| 93 | 2.213 ± 3.586 | 12.058 ± 9.917 | 56.018 ± 18.624 | 29.711 ± 18.287 | 177 |
| 94 | 2.574 ± 7.842 | 12.326 ± 10.560 | 57.750 ± 21.297 | 27.351 ± 18.621 | 177 |
| 95 | 2.611 ± 7.487 | 10.869 ± 11.220 | 55.446 ± 20.354 | 31.073 ± 19.047 | 205 |
| 96 | 1.465 ± 2.483 | 10.915 ± 9.385 | 58.132 ± 19.983 | 29.489 ± 19.539 | 207 |
| 97 | 2.873 ± 8.125 | 11.834 ± 10.494 | 56.371 ± 19.763 | 28.923 ± 17.779 | 189 |
| 98 | 1.723 ± 3.703 | 11.758 ± 11.691 | 57.077 ± 19.098 | 29.442 ± 17.300 | 178 |
| 99 | 2.369 ± 5.029 | 10.693 ± 9.744 | 58.345 ± 19.177 | 28.592 ± 17.301 | 157 |
| 100 | 2.252 ± 3.747 | 11.568 ± 10.480 | 55.997 ± 19.712 | 30.183 ± 18.918 | 181 |
| >101 | 2.293 ± 5.823 | 9.565 ± 9.766 | 57.883 ± 20.150 | 30.258 ± 17.890 | 7886 |
| | | | | | **Total**: 351849 |

**SUPPLEMENTARY TABLE 9.** Percentage of single-, first-, middle- and last-authorship papers in respect to the size of portfolio for research articles. As the number of available portfolios decreases as the size of portfolio increases the data has been bootstrapped 10,000 times (especially crucial for the portfolios with >50 items, where there is less than 1,000 of portfolios passing the threshold).

| Portfolio size | Single | First | Middle | Last | No. portfolios |
|---|---|---|---|---|---|
| 3 | 1.380 ± 8.283 | 34.260 ± 33.051 | 52.941 ± 34.279 | 11.419 ± 23.654 | 44969 |
| 4 | 1.194 ± 7.166 | 33.674 ± 29.977 | 54.326 ± 31.185 | 10.806 ± 21.446 | 33681 |
| 5 | 1.190 ± 6.721 | 32.690 ± 27.301 | 55.801 ± 28.793 | 10.319 ± 20.004 | 26474 |
| 6 | 1.050 ± 5.994 | 32.168 ± 25.798 | 56.757 ± 27.003 | 10.025 ± 18.904 | 22044 |
| 7 | 1.088 ± 5.892 | 31.458 ± 24.409 | 57.122 ± 26.088 | 10.331 ± 18.354 | 18366 |
| 8 | 1.091 ± 5.544 | 30.528 ± 23.273 | 57.853 ± 25.045 | 10.529 ± 17.980 | 15894 |
| 9 | 1.092 ± 5.639 | 30.162 ± 22.592 | 57.709 ± 24.383 | 11.037 ± 18.029 | 13644 |
| 10 | 0.900 ± 4.764 | 29.007 ± 21.600 | 58.518 ± 23.845 | 11.575 ± 18.141 | 12189 |
| 11 | 0.998 ± 5.178 | 28.859 ± 20.875 | 58.559 ± 23.113 | 11.584 ± 17.683 | 10502 |
| 12 | 1.013 ± 4.720 | 28.066 ± 20.541 | 58.709 ± 22.735 | 12.213 ± 17.712 | 9528 |
| 13 | 0.935 ± 4.474 | 27.175 ± 19.659 | 59.270 ± 22.280 | 12.620 ± 17.794 | 8647 |
| 14 | 0.939 ± 4.510 | 26.187 ± 19.080 | 59.335 ± 22.141 | 13.539 ± 18.230 | 7798 |
| 15 | 0.939 ± 4.524 | 25.676 ± 18.784 | 59.676 ± 21.906 | 13.709 ± 17.953 | 7228 |
| 16 | 0.934 ± 4.528 | 24.763 ± 18.391 | 59.692 ± 21.640 | 14.611 ± 18.551 | 6350 |
| 17 | 1.047 ± 5.222 | 24.708 ± 17.999 | 59.620 ± 21.731 | 14.624 ± 18.611 | 5934 |
| 18 | 0.872 ± 4.081 | 23.809 ± 17.650 | 59.409 ± 21.325 | 15.909 ± 18.841 | 5479 |
| 19 | 0.781 ± 3.718 | 23.155 ± 17.509 | 59.820 ± 21.460 | 16.244 ± 18.628 | 4992 |
| 20 | 0.798 ± 3.681 | 22.613 ± 17.073 | 59.856 ± 21.485 | 16.734 ± 19.026 | 4721 |
| 21 | 1.007 ± 4.440 | 22.195 ± 16.848 | 59.525 ± 21.153 | 17.272 ± 18.905 | 4255 |
| 22 | 1.037 ± 4.417 | 21.585 ± 16.549 | 59.745 ± 20.976 | 17.633 ± 18.936 | 3907 |
| 23 | 0.780 ± 3.390 | 20.918 ± 15.660 | 59.989 ± 20.416 | 18.313 ± 19.177 | 3718 |
| 24 | 0.825 ± 3.758 | 20.361 ± 15.753 | 59.776 ± 21.321 | 19.038 ± 19.749 | 3447 |
| 25 | 0.817 ± 3.535 | 20.423 ± 15.685 | 59.348 ± 20.385 | 19.412 ± 19.549 | 3252 |
| 26 | 0.823 ± 3.566 | 19.858 ± 15.579 | 59.656 ± 20.768 | 19.663 ± 19.353 | 2974 |
| 27 | 0.837 ± 3.389 | 19.707 ± 15.259 | 60.049 ± 20.443 | 19.407 ± 19.284 | 2689 |
| 28 | 0.847 ± 3.757 | 19.522 ± 15.294 | 59.747 ± 20.486 | 19.885 ± 18.817 | 2578 |
| 29 | 0.804 ± 2.975 | 18.969 ± 14.936 | 59.666 ± 20.249 | 20.561 ± 19.078 | 2442 |
| 30 | 0.804 ± 3.237 | 18.399 ± 15.049 | 58.998 ± 21.029 | 21.799 ± 20.121 | 2234 |
| 31 | 0.795 ± 3.992 | 18.275 ± 14.785 | 60.129 ± 20.624 | 20.801 ± 19.723 | 2171 |
| 32 | 0.793 ± 3.632 | 17.629 ± 14.525 | 59.301 ± 20.682 | 22.276 ± 19.986 | 1980 |
| 33 | 0.998 ± 4.098 | 17.819 ± 14.898 | 58.910 ± 20.638 | 22.273 ± 19.557 | 1861 |
| 34 | 0.752 ± 2.616 | 16.903 ± 13.841 | 59.441 ± 20.248 | 22.904 ± 19.627 | 1783 |
| 35 | 0.791 ± 2.761 | 17.363 ± 14.089 | 58.970 ± 19.909 | 22.876 ± 19.380 | 1697 |
| 36 | 0.843 ± 3.157 | 17.037 ± 14.214 | 58.169 ± 20.440 | 23.951 ± 20.197 | 1552 |
| 37 | 0.750 ± 3.422 | 15.701 ± 13.679 | 59.334 ± 20.200 | 24.215 ± 19.952 | 1586 |
| 38 | 0.718 ± 2.631 | 16.091 ± 13.549 | 58.903 ± 20.521 | 24.289 ± 20.183 | 1405 |
| 39 | 0.814 ± 3.360 | 15.664 ± 13.227 | 59.550 ± 20.078 | 23.972 ± 19.744 | 1303 |
| 40 | 0.686 ± 2.273 | 15.795 ± 13.456 | 59.847 ± 20.506 | 23.671 ± 19.594 | 1233 |
| 41 | 0.831 ± 3.193 | 15.473 ± 13.486 | 58.801 ± 20.478 | 24.895 ± 20.100 | 1207 |
| 42 | 0.751 ± 3.817 | 14.576 ± 12.733 | 60.057 ± 19.942 | 24.616 ± 19.912 | 1165 |
| 43 | 0.762 ± 2.439 | 15.430 ± 13.221 | 59.556 ± 20.352 | 24.252 ± 19.879 | 1108 |
| 44 | 0.687 ± 1.901 | 14.642 ± 13.166 | 59.259 ± 19.913 | 25.411 ± 19.689 | 1048 |
| 45 | 0.923 ± 4.101 | 14.603 ± 12.966 | 58.937 ± 20.414 | 25.538 ± 20.127 | 1014 |
| 46 | 0.801 ± 3.131 | 13.514 ± 12.028 | 60.824 ± 19.865 | 24.861 ± 19.094 | 939 |
| 47 | 0.743 ± 2.187 | 14.492 ± 13.046 | 58.906 ± 19.720 | 25.858 ± 19.684 | 930 |
| 48 | 0.849 ± 3.477 | 13.913 ± 13.105 | 59.010 ± 19.788 | 26.229 ± 19.483 | 827 |
| 49 | 0.753 ± 2.591 | 14.131 ± 12.877 | 60.214 ± 20.416 | 24.902 ± 19.415 | 772 |
| 50 | 0.909 ± 3.396 | 13.783 ± 12.561 | 59.228 ± 19.876 | 26.080 ± 19.957 | 774 |
| 51 | 0.835 ± 3.998 | 13.808 ± 12.211 | 58.557 ± 19.669 | 26.800 ± 19.015 | 702 |
| 52 | 0.853 ± 2.703 | 13.529 ± 12.358 | 59.052 ± 20.204 | 26.565 ± 20.275 | 715 |
| 53 | 0.655 ± 2.225 | 13.789 ± 12.249 | 58.539 ± 19.942 | 27.016 ± 19.158 | 656 |
| 54 | 0.836 ± 2.482 | 12.679 ± 11.868 | 58.645 ± 20.151 | 27.840 ± 20.131 | 688 |
| 55 | 0.648 ± 1.866 | 11.975 ± 11.085 | 59.745 ± 19.839 | 27.632 ± 19.849 | 616 |
| 56 | 0.971 ± 3.946 | 13.365 ± 13.641 | 58.799 ± 20.030 | 26.865 ± 19.565 | 616 |
| 57 | 1.177 ± 5.208 | 12.247 ± 12.119 | 59.518 ± 20.079 | 27.058 ± 19.349 | 615 |
| 58 | 0.617 ± 3.432 | 10.347 ± 11.880 | 65.948 ± 23.328 | 23.088 ± 20.562 | 640 |
| 59 | 0.780 ± 2.596 | 12.534 ± 11.795 | 59.886 ± 20.541 | 26.799 ± 19.484 | 552 |
| 60 | 0.763 ± 2.786 | 12.278 ± 11.559 | 58.173 ± 19.616 | 28.786 ± 19.499 | 559 |
| 61 | 0.689 ± 2.331 | 11.254 ± 11.263 | 59.291 ± 19.111 | 28.766 ± 19.487 | 463 |
| 62 | 0.610 ± 1.663 | 12.005 ± 11.923 | 58.757 ± 19.578 | 28.629 ± 19.416 | 447 |
| 63 | 0.750 ± 2.379 | 11.751 ± 11.751 | 59.358 ± 19.507 | 28.141 ± 19.881 | 424 |
| 64 | 0.613 ± 1.822 | 11.784 ± 11.490 | 58.273 ± 20.658 | 29.330 ± 20.369 | 429 |
| 65 | 0.652 ± 1.446 | 12.257 ± 12.254 | 57.313 ± 19.124 | 29.779 ± 19.361 | 401 |
| 66 | 0.732 ± 2.138 | 11.831 ± 11.231 | 59.378 ± 18.293 | 28.060 ± 18.161 | 429 |
| 67 | 0.813 ± 2.362 | 11.519 ± 10.763 | 57.974 ± 18.953 | 29.694 ± 18.606 | 386 |
| 68 | 0.986 ± 4.039 | 11.317 ± 12.043 | 60.323 ± 20.463 | 27.374 ± 19.186 | 383 |
| 69 | 0.711 ± 1.733 | 11.889 ± 11.604 | 59.726 ± 20.278 | 27.675 ± 19.361 | 346 |
| 70 | 0.915 ± 4.095 | 11.176 ± 9.823 | 58.476 ± 18.274 | 29.433 ± 18.873 | 350 |
| 71 | 0.581 ± 1.559 | 11.211 ± 11.290 | 60.264 ± 19.644 | 27.945 ± 19.235 | 371 |
| 72 | 0.597 ± 1.428 | 10.572 ± 11.089 | 59.816 ± 19.944 | 29.016 ± 19.653 | 303 |
| 73 | 1.285 ± 6.514 | 11.571 ± 11.709 | 56.214 ± 20.008 | 30.931 ± 19.798 | 323 |
| 74 | 0.710 ± 2.612 | 10.699 ± 10.892 | 58.622 ± 19.189 | 29.970 ± 18.964 | 316 |
| 75 | 0.990 ± 2.595 | 10.047 ± 10.631 | 61.387 ± 20.152 | 27.576 ± 19.779 | 292 |
| 76 | 1.046 ± 4.652 | 11.238 ± 11.583 | 58.666 ± 20.003 | 29.050 ± 19.957 | 294 |
| 77 | 0.716 ± 2.580 | 10.629 ± 11.040 | 59.610 ± 20.770 | 29.046 ± 20.478 | 301 |
| 78 | 0.610 ± 1.849 | 9.325 ± 9.723 | 61.094 ± 19.908 | 28.971 ± 18.368 | 261 |
| 79 | 0.909 ± 2.570 | 9.528 ± 9.396 | 58.509 ± 19.882 | 31.054 ± 19.028 | 260 |
| 80 | 0.598 ± 1.293 | 9.263 ± 8.634 | 60.192 ± 18.345 | 29.948 ± 18.184 | 251 |
| 81 | 0.905 ± 3.267 | 9.458 ± 10.038 | 60.201 ± 21.051 | 29.436 ± 19.844 | 238 |
| 82 | 0.578 ± 1.836 | 10.464 ± 10.024 | 60.869 ± 18.375 | 28.088 ± 18.615 | 232 |
| 83 | 0.643 ± 1.441 | 10.352 ± 9.920 | 60.152 ± 18.069 | 28.853 ± 17.875 | 212 |
| 84 | 0.861 ± 2.705 | 9.442 ± 9.311 | 57.823 ± 17.191 | 31.875 ± 17.827 | 210 |
| 85 | 0.555 ± 1.399 | 9.657 ± 9.854 | 58.577 ± 17.922 | 31.211 ± 19.017 | 226 |
| 86 | 0.813 ± 3.363 | 9.772 ± 10.393 | 58.877 ± 19.054 | 30.539 ± 19.014 | 196 |
| 87 | 1.073 ± 3.856 | 9.510 ± 10.790 | 57.054 ± 20.211 | 32.363 ± 20.356 | 188 |
| 88 | 0.587 ± 1.701 | 10.994 ± 12.358 | 58.868 ± 20.416 | 29.551 ± 19.698 | 180 |
| 89 | 1.163 ± 4.010 | 8.492 ± 9.073 | 60.683 ± 18.656 | 29.661 ± 17.923 | 176 |
| 90 | 0.606 ± 1.298 | 9.532 ± 9.371 | 58.402 ± 18.139 | 31.459 ± 18.583 | 144 |
| 91 | 0.502 ± 1.386 | 8.379 ± 9.133 | 58.610 ± 20.110 | 32.508 ± 19.768 | 170 |
| 92 | 1.185 ± 6.106 | 8.558 ± 8.160 | 59.011 ± 20.369 | 31.246 ± 20.021 | 171 |
| 93 | 0.513 ± 1.049 | 8.568 ± 8.092 | 61.236 ± 17.628 | 29.682 ± 18.231 | 144 |
| 94 | 0.627 ± 1.439 | 8.993 ± 8.827 | 61.518 ± 18.521 | 28.862 ± 17.611 | 158 |
| 95 | 0.508 ± 1.091 | 9.190 ± 10.099 | 64.794 ± 19.374 | 25.508 ± 19.080 | 143 |
| 96 | 0.582 ± 1.409 | 8.927 ± 11.546 | 59.708 ± 20.188 | 30.783 ± 18.316 | 159 |
| 97 | 0.758 ± 1.473 | 8.880 ± 8.247 | 62.519 ± 17.751 | 27.844 ± 16.956 | 136 |
| 98 | 0.562 ± 1.834 | 8.543 ± 9.439 | 60.384 ± 18.139 | 30.511 ± 18.632 | 143 |
| 99 | 0.964 ± 2.528 | 8.635 ± 9.823 | 60.370 ± 21.707 | 30.030 ± 20.460 | 129 |
| 100 | 0.536 ± 1.226 | 7.586 ± 7.629 | 60.952 ± 19.619 | 30.926 ± 19.352 | 142 |
| >100 | 0.638 ± 2.156 | 7.161 ± 8.847 | 62.791 ± 20.440 | 29.409 ± 18.944 | 5712 |
| | | | | | **Total**: 338889 |

**SUPPLEMENTARY TABLE 10.** A selection of bibliometric statistics for some prominent scientists. For the screenshots from fCite see http://www.fcite.org/examples.html

| Name | Field | In HCR | Articles Number | Avg No Authors | Single % | First % | Middle % | Last % | fCite | | | | | | | Google Scholar | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | $FLAE_{RCR}$ | $EC_{RCR}$ | Total RCR | FLAE/RCR | $FLAE_{Cit}$ | $EC_{Cit}$ | Citations | H-index | Citations |
| Scientist A (ro) | Clinical Medicine | 1 | 132 | 12.7 | 0.00 | 5.30 | 10.61 | 84.09 | 35.5 | 56.7 | 803.4 | 4.4 % | 1017 | 1604 | 21081 | - | - |
| Scientist A (all) | | | 163 | 15.8 | 0.61 | 4.29 | 85.28 | 9.82 | 44.2 | 69.7 | 1239.9 | 3.6 % | 1248 | 1952 | 32061 | | |
| Scientist B (ro) | Physics | 0 | 544 | 1379.3 | 0.18 | 0.37 | 0.18 | 99.26 | 1.3 | 2.0 | 335.6 | 0.4 % | 26 | 43 | 2661 | 191 | 283,988 |
| Scientist B (all) | | | 548 | 1369.2 | 0.36 | 0.36 | 0.36 | 98.91 | 2.5 | 2.8 | 338.0 | 0.7 % | 54 | 61 | 2717 | | |
| Scientist C (ro) | Molecular Biology & Genetics | 0 | 33 | 15.3 | 36.36 | 21.21 | 36.36 | 6.06 | 1087.8 | 1064.7 | 1286.7 | 84.5 % | 28429 | 27311 | 32268 | 34 | 60,753 |
| Scientist C (all) | | | 35 | 14.9 | 34.29 | 22.86 | 37.14 | 5.71 | 1090.1 | 1025.8 | 1290.4 | 84.5 % | 28510 | 27372 | 32391 | | |
| Scientist D (ro) | Physics | 0 | 174 | 2208.9 | 0.00 | 0.0 | 100.00 | 0.00 | 0.05 | 0.05 | 114.3 | 0.0 % | 0.3 | 0.3 | 697 | 168 | 127,547 |
| Scientist D (all) | | | 174 | 2208.9 | 0.00 | 0.0 | 100.00 | 0.00 | 0.05 | 0.05 | 114.3 | 0.0 % | 0.3 | 0.3 | 697 | | |
| Scientist E (ro) | Materials Science | 1 | 108 | 10.2 | 0.00 | 5.56 | 65.74 | 28.70 | 100.0 | 67.8 | 512.7 | 19.3 % | 2323 | 1497 | 11436 | 107 | 99,769 |
| Scientist E (all) | | | 112 | 10.0 | 0.00 | 5.36 | 65.18 | 29.46 | 103.9 | 74.8 | 551.7 | 18.8 % | 2400 | 1605 | 12037 | | |
| Scientist F (ro) | Clinical Medicine | 1 | 156 | 32.4 | 10.26 | 10.90 | 53.85 | 25.00 | 69.5 | 78.6 | 2305.8 | 3.0 % | 1606 | 1690 | 31927 | 102 | 104,919 |
| Scientist F (all) | | | 296 | 20.6 | 14.19 | 15.54 | 44.93 | 25.34 | 121.1 | 140.2 | 2769.5 | 4.4 % | 2402 | 2695 | 39986 | | |
| Scientist G (ro) | Physics | 1 | 237 | 8.1 | 0.42 | 2.38 | 62.03 | 34.18 | 93.4 | 113.9 | 924.9 | 10.1 % | 1694 | 1820 | 14612 | 152 | 113,094 |
| Scientist G (all) | | | 246 | 8.0 | 1.22 | 3.25 | 60.57 | 34.96 | 95.8 | 115.7 | 934.7 | 10.2 % | 1724 | 1845 | 14735 | | |
| Scientist H (ro) | Molecular Biology & Genetics | 1 | 185 | 30.0 | 0.54 | 0.00 | 64.86 | 34.59 | 391.4 | 484.8 | 2949.4 | 13.3 % | 12021 | 14808 | 90148 | 128 | 173,483 |
| Scientist H (all) | | | 197 | 30.4 | 0.51 | 0.51 | 65.48 | 33.50 | 394.6 | 489.5 | 2982.0 | 13.2 % | 12111 | 14939 | 90864 | | |
| Scientist I (ro) | Biology & Biochemistry | 1 | 75 | 19.3 | 0.00 | 14.67 | 85.33 | 0.00 | 106.1 | 84.7 | 939.3 | 11.3 % | 2283 | 1791 | 20596 | 71 | 58,008 |
| Scientist I (all) | | | 83 | 18.3 | 1.20 | 16.87 | 80.72 | 1.20 | 110.0 | 87.6 | 948.2 | 11.6 % | 2372 | 1857 | 20816 | | |
| Scientist J (ro) | Biology & Biochemistry | 1 | 296 | 9.1 | 0.00 | 0.68 | 63.18 | 36.15 | 145.9 | 182.9 | 856.8 | 17.0 % | 4172 | 5100 | 23419 | 115 | 68,249 |
| Scientist J (all) | | | 350 | 8.8 | 2.00 | 1.43 | 60.86 | 35.71 | 183.4 | 222.7 | 1022.6 | 17.9 % | 5520 | 6490 | 29112 | | |
| Scientist K (ro) | Chemistry | 1 | 286 | 7.0 | 0.35 | 0.70 | 59.09 | 39.86 | 111.2 | 132.3 | 748.5 | 14.9 % | 2177 | 2504 | 12871 | 128 | 67,952 |
| Scientist K (all) | | | 297 | 6.9 | 0.34 | 1.01 | 59.26 | 39.39 | 131.7 | 149.2 | 848.7 | 15.5 % | 2523 | 2769 | 14407 | | |
| Scientist L (ro) | Chemistry | 1 | 194 | 8.2 | 0.00 | 2.00 | 55.15 | 42.78 | 65.3 | 56.4 | 394.1 | 16.6 % | 1197 | 1058 | 7313 | 87 | 37,474 |
| Scientist L (all) | | | 206 | 8.0 | 0.00 | 1.94 | 55.34 | 42.72 | 73.5 | 67.8 | 448.6 | 16.4 % | 1284 | 1203 | 7984 | | |
| Scientist M (ro) | Immunology | 1 | 187 | 10.1 | 1.07 | 0.00 | 66.84 | 32.09 | 139.7 | 163.8 | 1011.0 | 13.8 % | 5321 | 6451 | 34327 | 135 | 97,049 |
| Scientist M (all) | | | 254 | 8.4 | 3.94 | 5.91 | 54.72 | 35.43 | 226.9 | 256.3 | 1233.2 | 18.4 % | 9036 | 10199 | 42998 | | |
| Scientist N (ro) | Immunology | 1 | 470 | 17.5 | 0.64 | 1.27 | 74.73 | 23.14 | 79.7 | 68.4 | 1082.3 | 7.4 % | 2492 | 1945 | 28841 | 132 | 65,087 |
| Scientist N (all) | | | 629 | 14.6 | 1.59 | 3.81 | 64.44 | 30.00 | 153.1 | 140.2 | 1368.9 | 6.9 % | 4639 | 3892 | 36113 | | |
| Scientist O (ro) | Microbiology | 1 | 243 | 9.8 | 0.00 | 1.65 | 65.02 | 33.33 | 100.9 | 99.7 | 727.0 | 13.9 % | 3085 | 3127 | 21307 | 108 | 44,840 |
| Scientist O (all) | | | 310 | 8.4 | 2.26 | 5.48 | 53.87 | 38.39 | 165.5 | 161.6 | 884.9 | 18.7 % | 4665 | 4558 | 24533 | | |
| Scientist P (ro) | Microbiology | 1 | 485 | 15.1 | 0.41 | 1.44 | 86.80 | 11.34 | 141.9 | 168.4 | 2488.7 | 5.7 % | 3183 | 3318 | 51992 | 135 | 93,706 |
| Scientist P (all) | | | 558 | 14.4 | 1.25 | 1.79 | 80.65 | 16.31 | 152.1 | 182.1 | 2562.0 | 5.9 % | 3433 | 3635 | 53533 | | |
| Scientist Q (ro) | Plant & Animal Science | 1 | 164 | 3.8 | 3.05 | 9.15 | 22.56 | 65.24 | 161.6 | 178.2 | 529.0 | 30.6 % | 4094 | 4502 | 13216 | 107 | 38,678 |
| Scientist Q (all) | | | 191 | 3.6 | 5.76 | 9.95 | 19.90 | 64.40 | 240.8 | 267.1 | 711.1 | 33.9 % | 6192 | 6727 | 17729 | | |
| Scientist R (ro) | Plant & Animal Science | 1 | 135 | 8.0 | 0.74 | 4.44 | 60.00 | 34.81 | 35.9 | 38.9 | 249.4 | 14.4 % | 779 | 811 | 4771 | 66 | 12,896 |
| Scientist R (all) | | | 177 | 7.1 | 1.13 | 7.91 | 50.85 | 40.11 | 55.9 | 60.7 | 313.4 | 17.8 % | 1211 | 1261 | 5909 | | |

***In HCR*** *– means that the researcher is included among Highly Cited Researchers 2018 list by Clarivate Analytics*
***ro/all*** *– **r**esearch **o**nly or **all** articles*