**ORIGINAL PAPER**

# Contemporary scientometric analyses using a novel web application: the science performance evaluation (SciPE) approach

Sean S. Scholz[1] · Markus Dillmann[2] · Alexander Flohr[2] · Christina Backes[2] · Tobias Fehlmann[2] · Dominic Millenaar[1] · Christian Ukena[1] · Michael Böhm[1] · Andreas Keller[2,3] · Felix Mahfoud[1,4]

## Abstract

**Aims**  We aimed at developing a structured study protocol utilizing the bibliographic web-application science performance evaluation (SciPE) to perform comprehensive scientometric analyses.

**Methods and results**  Metadata related to publications derived from online databases were processed and visualized by transferring the information to an undirected multipartite graph and distinct partitioned sets of nodes. Also, institution-specific data were normalized and merged allowing precise geocoordinate positioning, to enable heatmapping and valid identification. As a result, verified, processed data regarding articles, institutions, journals, authors gender, nations and subject categories can be obtained. We recommend including the total number of publications, citations, the population, research institutions, gross domestic product, and the country-specific modified Hirsch Index and to form corresponding ratios (e.g., population/publication). Also, our approach includes implementation of bioinformatical methods such as heatmapping based on exact geocoordinates, simple chord diagrams, and the central implementation of specific ratios with plain visualization techniques.

**Conclusion**  This protocol allows precise conduction of contemporaneous scientometric analyses based on bioinformatic and meta-analytical techniques, allowing to evaluate and contextualize scientific efforts. Data presentation with the depicted visualization techniques is mandatory for transparent and consistent analyses of research output across different nations and topics. Research performance can then be discussed in a synopsis of all findings.

Andreas Keller and Felix Mahfoud contributed equally to this work.

✉ Sean S. Scholz
sean.scholz@uks.eu

1  Klinik für Innere Medizin III, Kardiologie, Angiologie und Internistische Intensivmedizin, Saarland University Hospital, Kirrberger Str. 1, Geb. 41, IMED, 66421 Homburg/Saar, Germany

2  Zentrum für Bioinformatik Saar, Abteilung für klinische Bioinformatik, Universität des Saarlandes, Saarbrücken, Germany

3  Department of Neurology, Stanford University California, California, CA, USA

4  Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA

**Graphic abstract**
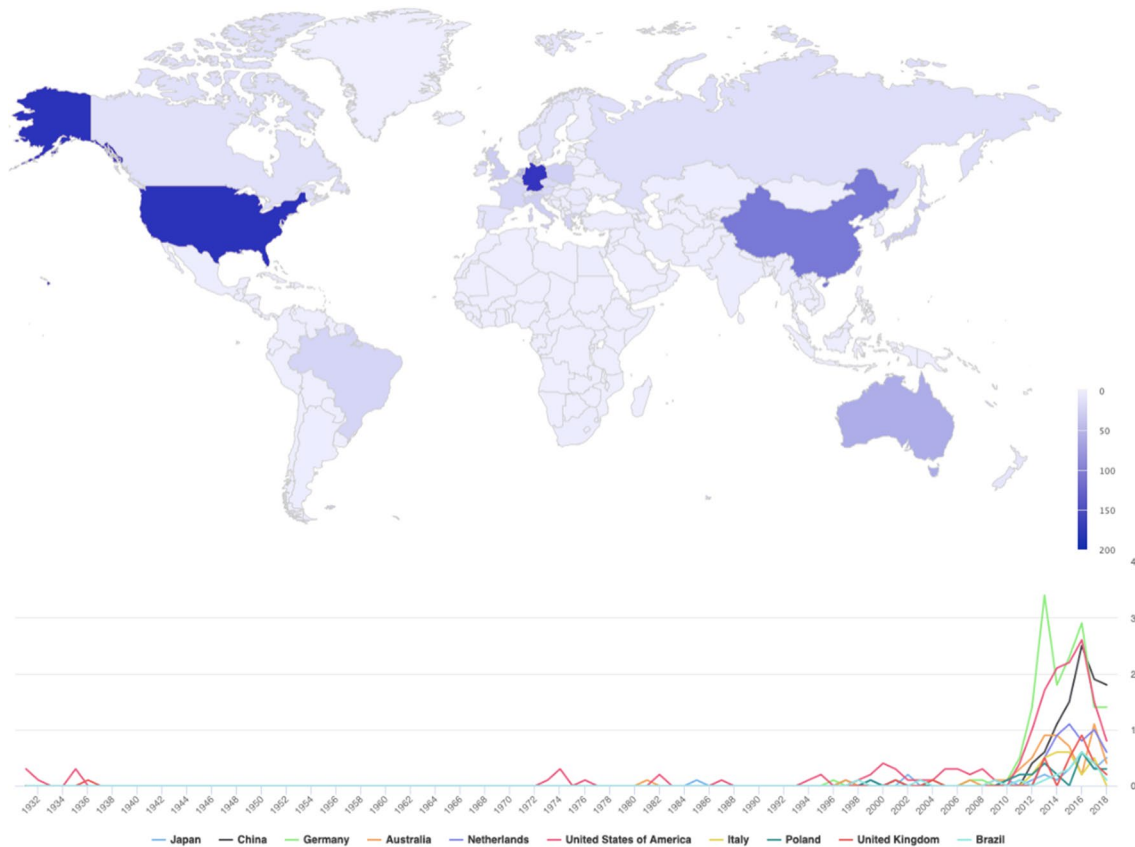


**Keywords**  Scientometry · Bibliometry · Bibliometrics · Meta-analysis · Research assessment

## Introduction

Global research output has grown exponentially over the past decades [1]. This development can be observed across various research areas and raises various questions, such as how the quality of scientific output can be objectively assessed. Many databases which provide research articles simultaneously collect metadata, e.g., information about author, affiliation, institute, subject area, date, journal, and the number of citations [2, 3]. This information can be utilized with computer science-driven techniques, allowing a broad possibility of analyses though limited to the underlying metadata. Several approaches for scientometric analyses have been presented [4–8]. However, implementation of bioinformatical methods including plain visualization of international research networks/distribution, such as heatmapping based on exact geocoordinates, and central implementation of tables including specific ratios, are, to the best of our knowledge, lacking. Data presentation with visualization (e.g., heatmaps, chord diagrams) techniques is mandatory for transparent and consistent analyses of research output

across nations and topics. Hence, such approach should allow evaluate and contextualize scientific efforts, which is becoming increasingly important in several research areas [9]. In this sense, apart from quantitative aspects (number of publications/citations), semiqualitative aspects of research, e.g., citation rates are of interest and should be taken into account. It is important that if research activities are analyzed globally, even a mere nationwide review of research activities is non-differentiated, though, individual institutions can provide a substantial contribution to a research area. Consequently, our approach not only includes nations and institutions, but also individual researchers allowing gender-specific research which itself is of growing interest [10, 11].

## The SciPE software

Science performance evaluation (SciPE) is a scientometric web application for publication analyses developed by the Institute of Bioinformatics and the Clinic of Internal

Medicine III, Saarland University, Saarbrücken and Homburg/Saar, Germany, which implements bioinformatic methods and visualization techniques to scientometric objectives. It is planned to make the web application publicly available.

## Data search and data extraction

The creation of a specific search term is of considerable interest and should be performed analogous to meta-analyses [12–20]. Ideally, besides including abbreviations and different spellings, multiple languages should be systematically included, as shown here in an exemplary search term aiming to identify articles related to percutaneous coronary interventions [TI = (percutaneous coronary Intervent* OR percutaneous coronary angioplasty OR coronary angiography OR Drug eluting stent OR Drug coated balloon OR balloon angioplasty OR coronary stent Implantation OR Cardio Angiographie OR Koronarangiographie OR cardiac catheterizatio* OR Herz* AND (katheter* OR Sond*) OR PTCA OR PCI)]. Subsequently, a total of 23,943 articles for the period between 1900 and 2018 were identified. We suggest inclusion of original articles only (exclusion of non-original articles, e.g., reviews and case reports) to improve quality and comparability, whereas title search only increases the representativeness of the database. The term can be searched using different databases such as Web of Science (WoS) and PubMed [3, 21]. In this context, SciPE provides two distinct not combinable data acquisition methods: first, data pre-selection with the online web application of WoS by Clarivate analytics [3]; second, data acquisition method can be performed by entering certain search termini, which are internally translated and passed to PubMed [21], allowing SciPE to gain access to all publicly available information on scientific publications allowing a similar analysis, however, with the lack of citation metrics provided by WoS.

## Data verification

After data extraction, a second step of data processing was implemented to guarantee accuracy. Besides internal processing, the complete analytical processes and all underlying information can be validated using the data tables section of SciPE to ensure transparency. As a result, valid assignment of publication data, e.g., country (based on the first authors affiliation), authorship, journals, articles, topics, and date are essential elements provided by SciPE. Based on the users input method, data are extracted and used to provide an undirected multipartite graph with distinct partitioned sets of nodes, e.g., for publications, institutes, authors, and subject categories. Apart from these directly derivable properties, SciPE implements specific analyses concerning institutions and gender of contributing authors as stated below.

## Analysis of institutes

Precise identification and localization of institutions are of special interest and consequently analysis of institutions is a key element of this protocol. Therefore, institute-specific data are analyzed in two steps. First, all institute names are compared to a normalized list of universities, which are internally compared by SciPE to a publicly available online university ranking list [22]. Afterward, those normalized university names are geocoded using the chargeable Google-Geocoding service. With this service, the exact geo-position of an institution with longitude and latitude can be derived and used to validate individual institute's positions and enables the possibility to create research-specific heatmaps. This geocoordinate merging of affiliated institutions with different names is not only necessary when campus-structured universities are analyzed, but also a promising attempt at the systematic improvement of scientometric research. However, this feature is coupled to an existing fee-based Google API-key which must be provided by the user. As a result, research related to the search term can be visualized globally (Fig. 1). Also, whenever SciPE is used, the institutes in the query will be geocoded and the distance to an institute of the normalized institution list noted. All corresponding departments, which are related to an institution within a range of 20 km, are automatically processed to the institutional name provided by the normalized set. All remaining institutions are internally resolved using the given information, the query itself, and their similarity.

## Analysis of research performance over time

After data verification, the processed data can be used to get basic information on the related research article. For example, medical publications can be assessed over a specific period of time, the number of authors, gender, and the proportion of each nation. Therefore, trends in different research areas as well as the gender distribution can be portrayed (Fig. 2).

## Resolving authorship and gender

Gender assignment is determined by SciPE using an open source python library [23]. According to this list, every author is allocated to one of the following gender categories male, female, and unknown in a nation-specific manner. Also, manual correction of the allocated gender is provided, if comprehensive gender-specific analyses are desired.

## Visualization of research, networks, and performance

Research collaborations are suspected to be a crucial element not only in global, but also in regional and local
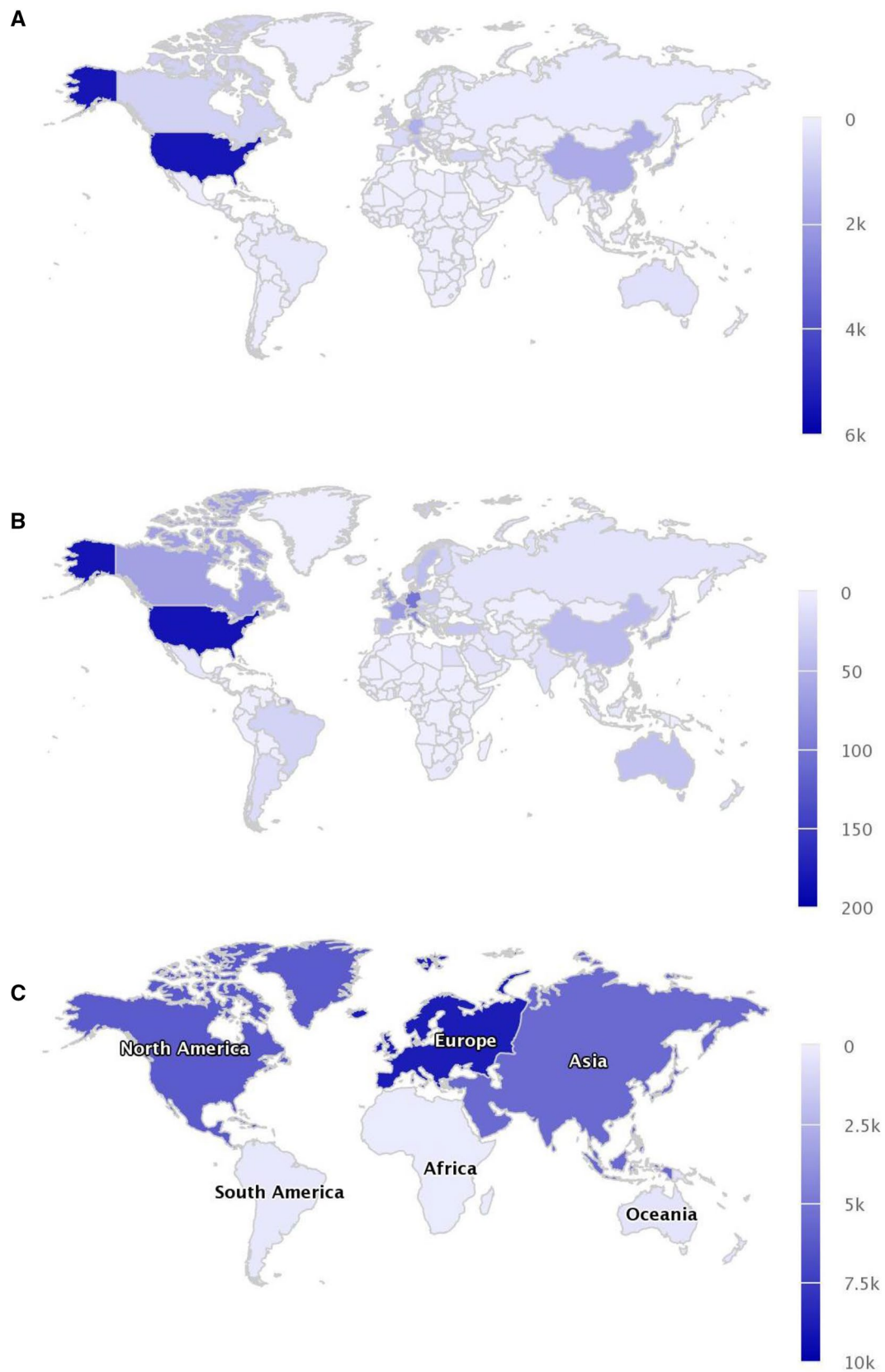
**Fig. 1** World map illustrating global research activities on percutaneous coronary interventions (1900–2018); **a** number of publications per nation, **b** H-Index per nation, **c** number of publications per continent; $k = 1000$ (query date: 20.10.19)
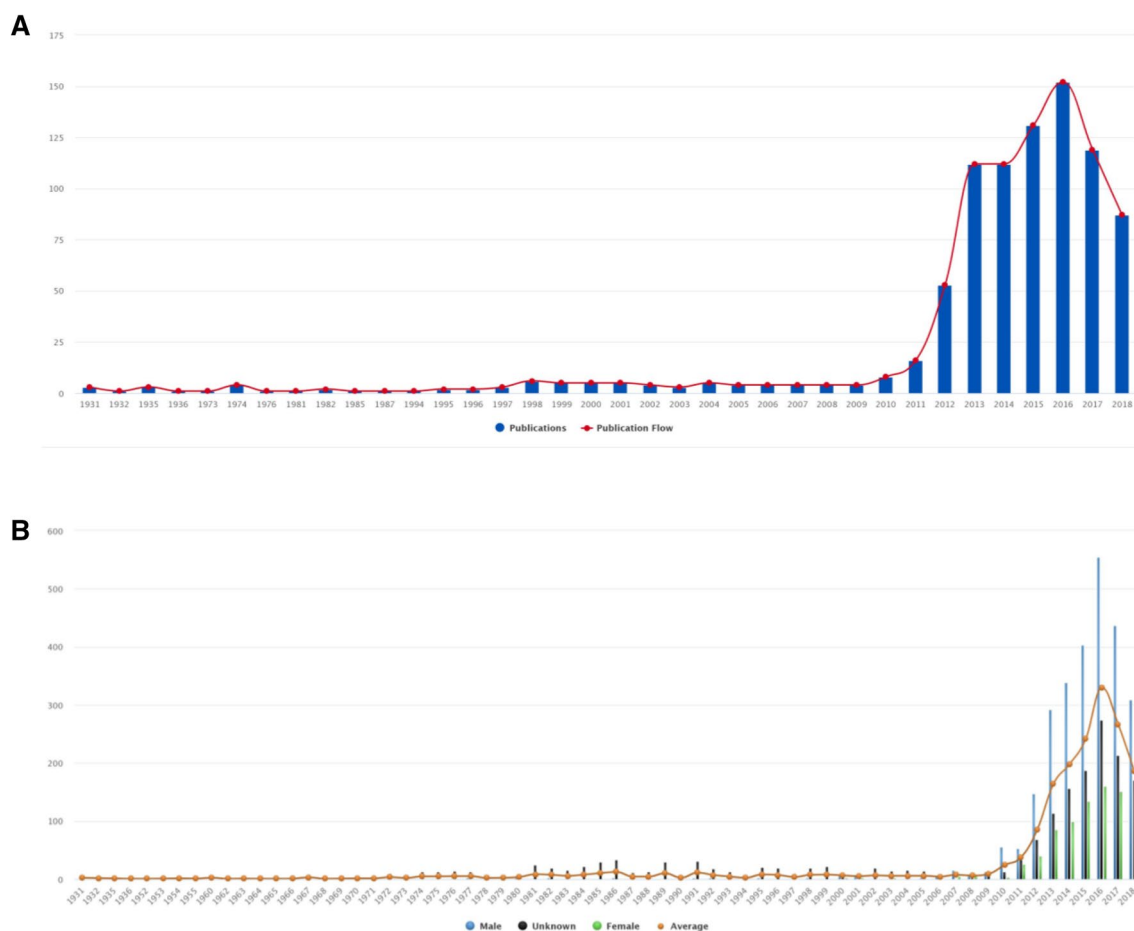
**Fig. 2** Global research activities on renal denervation (1931–2018): **a** number of publications per year, **b** number of authors and gender distribution per year (query date: 20.10.19)

research clusters [9]. Moreover, simple and correct visualization of these network patterns is of special interest to fully understand research architecture. SciPE primarily uses chord diagrams to visualize this cooperation between authors, institutions, and nations. In these chord diagrams, the number of joint publications are visualized as proportional links between cooperating institutes. This is visualized by an exemplary search for renal denervation that shows interinstitutional collaborations (Fig. 3). Consequently, these diagrams can be used to identify well-connected nations, authors, and institutions and, under consideration of overall research performance, how collaborations influence research effectiveness.

## Evaluation of research performance

Besides quantification of research, qualitative aspects have to be analyzed and evaluated carefully. For this reason, scientometric indexes such as the Hirsch Index (H-Index) can be incorporated into the analysis, and expanded to nations, institutes, and gender categories [24]. As this index relies

on citation rates which are not provided by PubMed, this option is only available in WoS. Aside from using the H-Index, the total number of institutions, authors, articles, and citations has to be taken into account. Furthermore, the financial power of nations, as well as the population, has to be considered when research performance is assessed. In this context, comprehensive evaluation of research efforts has to include at least the total number of publications, citations, the population, gross domestic product (GDP), and the country-specific H-Index in tabular form, as presented in Table 1. We recommend including all mentioned parameters and additionally to form corresponding ratios (e.g., population per publication/H-Index/citation) as visualized in Table 2. If these semiqualitative criteria are included in the analysis, it is possible to make a differentiated and easily objectifiable statement regarding research performance. We recommend visualizing this in at least one table to provide a simple overview of the included aspects. The overall evaluation of research performance can then be discussed in a synopsis of all findings.
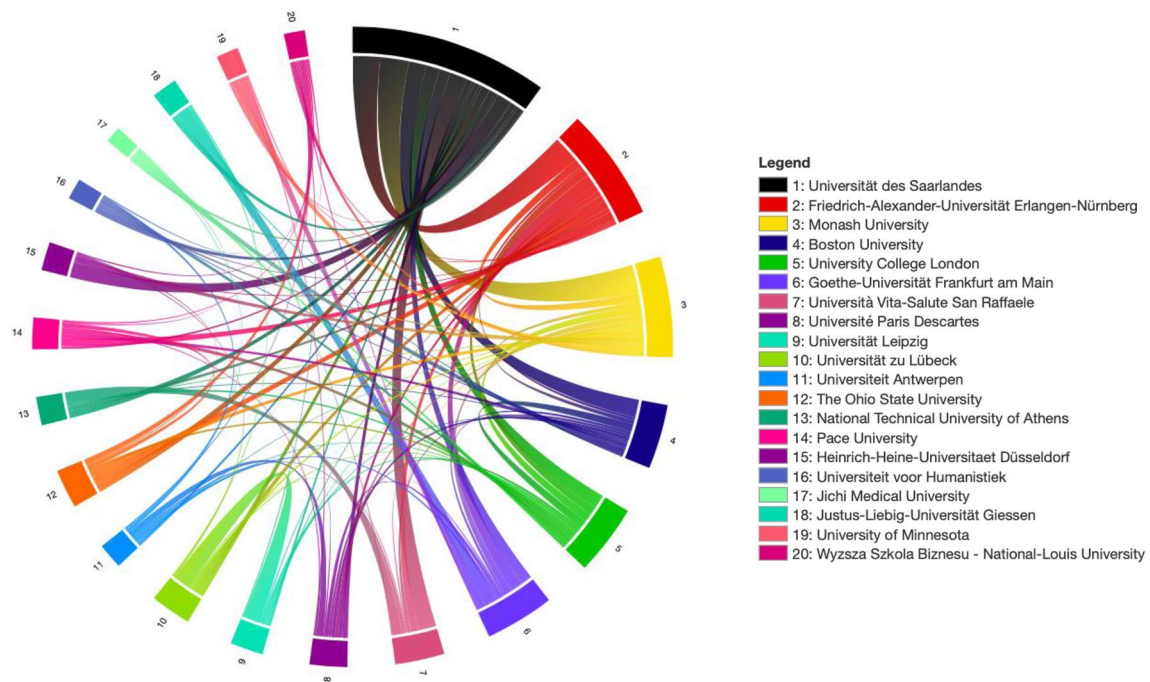
**Legend**
1: Universität des Saarlandes
2: Friedrich-Alexander-Universität Erlangen-Nürnberg
3: Monash University
4: Boston University
5: University College London
6: Goethe-Universität Frankfurt am Main
7: Università Vita-Salute San Raffaele
8: Université Paris Descartes
9: Universität Leipzig
10: Universität zu Lübeck
11: Universiteit Antwerpen
12: The Ohio State University
13: National Technical University of Athens
14: Pace University
15: Heinrich-Heine-Universitaet Düsseldorf
16: Universiteit voor Humanistiek
17: Jichi Medical University
18: Justus-Liebig-Universität Giessen
19: University of Minnesota
20: Wyzsza Szkola Biznesu - National-Louis University

**Fig. 3** Chord diagram illustrating institutional research collaborations on renal denervation (1931–2018); width of connection strand between institutions is proportional to extent of cooperation between the two (query date: 20.10.19)

**Table 1** Top 20 nations regarding publications on percutaneous coronary interventions (1900–2018), ranked by modified H-Index

|   | Country | H-Index | Articles | Citations | Population [27] | GDP (in million) [27] |
|---|---------|---------|----------|-----------|-----------------|------------------------|
| 1 | USA | 183 | 5439 | 203,307 | 329,256,465 | 19,490,000 |
| 2 | Germany | 101 | 1592 | 45,812 | 80,457,737 | 4,199,000 |
| 3 | Netherlands | 96 | 1033 | 41,147 | 17,151,228 | 924,400 |
| 4 | Italy | 82 | 1234 | 31,269 | 62,246,674 | 2,317,000 |
| 5 | France | 68 | 524 | 16,862 | 67,364,357 | 2,856,000 |
| 6 | Japan | 66 | 1654 | 25,871 | 126,168,156 | 5,443,000 |
| 7 | UK | 66 | 852 | 20,172 | 65,105,246 | 2,925,000 |
| 8 | Canada | 65 | 707 | 21,188 | 35,881,659 | 1,774,000 |
| 9 | Switzerland | 60 | 388 | 15,958 | 8,292,809 | 523,100 |
| 10 | South Korea | 56 | 989 | 16,367 | 51,418,097 | 2,035,000 |
| 11 | China | 42 | 1689 | 13,816 | 1,384,688,986 | 23,210,000 |
| 11 | Israel | 40 | 249 | 5814 | 8,424,904 | 317,100 |
| 13 | Sweden | 40 | 232 | 6291 | 10,040,995 | 518,000 |
| 14 | Spain | 38 | 335 | 7309 | 49,331,076 | 1,778,000 |
| 15 | Denmark | 38 | 298 | 6678 | 5,809,502 | 287,800 |
| 16 | Australia | 37 | 363 | 5383 | 23,470,145 | 1,248,000 |
| 17 | Belgium | 37 | 196 | 6153 | 11,570,762 | 529,200 |
| 18 | Austria | 34 | 198 | 4814 | 8,793,370 | 441,000 |
| 19 | Turkey | 32 | 635 | 5552 | 78,741,053 | 2,186,000 |
| 20 | Greece | 32 | 191 | 3520 | 10,858,018 | 299,300 |

*GDP* Gross domestic product, *USA* United States of America, *UK* United Kingdom (Query date: 20.10.19)

**Table 2** Research on percutaneous coronary intervention (1900–2018) ranked by scientific output

| Rank | Population per… | | | Publications per… | Citations per… |
|---|---|---|---|---|---|
| | (A) Articles | (B) H-Index | (C) Citations | (D) GDP | (E) Articles |
| 1 | Netherlands 16603,31 | Switzerland 138213,48 | Netherlands 416,83 | Netherlands 0,00112 | Switzerland 41,13 |
| 2 | Denmark 19494,97 | Denmark 152881,63 | Switzerland 519,66 | Denmark 0,00104 | Netherlands 39,83 |
| 3 | Switzerland 21373,22 | Netherlands 178658,63 | Denmark 869,95 | Israel 0,00079 | USA 37,38 |
| 4 | Israel 33834,96 | Israel 210622,60 | Israel 1449,07 | Switzerland 0,00074 | France 32,18 |
| 5 | Sweden 43280,15 | Sweden 251024,88 | Sweden 1596,09 | Greece 0,00064 | Belgium 31,39 |
| 6 | Austria 44410,96 | Austria 258628,53 | USA 1619,50 | Italy 0,00053 | Canada 29,97 |
| 7 | Italy 50443,01 | Belgium 312723,30 | Canada 1693,49 | South Korea 0,00049 | Germany 28,78 |
| 8 | Germany 50538,78 | Greece 339313,06 | Germany 1756,26 | Sweden 0,00045 | Sweden 27,12 |
| 9 | Canada 50751,99 | Canada 552025,52 | Austria 1826,63 | Austria 0,00045 | Italy 25,34 |
| 10 | South Korea 51989,98 | Australia 634328,24 | Belgium 1880,51 | Canada 0,0004 | Austria 24,31 |
| 11 | Greece 56848,26 | Italy 759105,78 | Italy 1990,68 | Germany 0,00038 | UK 23,68 |
| 12 | Belgium 59034,50 | Germany 796611,26 | Greece 3084,66 | Belgium 0,00037 | Israel 23,35 |
| 13 | USA 60536,21 | South Korea 918180,30 | South Korea 3141,57 | Japan 0,0003 | Denmark 22,41 |
| 14 | Australia 64656,05 | UK 986443,12 | UK 3227,51 | United Kingdom 0,00029 | Spain 21,82 |
| 15 | Japan 76280,63 | France 990652,31 | France 3995,04 | Australia 0,00029 | Greece 18,43 |
| 16 | UK 76414,61 | Spain 1298186,20 | Australia 4360,05 | Turkey 0,00029 | South Korea 16,55 |
| 17 | Turkey 124001,66 | USA 1799215,70 | Japan 4876,82 | USA 0,00028 | Japan 15,64 |
| 18 | France 128557,93 | Japan 1911638,70 | Spain 6749,36 | Spain 0,00019 | Australia 14,83 |
| 19 | Spain 147256,94 | Turkey 2460657,90 | Turkey 14182,47 | France 0,00018 | Turkey 8,74 |
| 20 | China 819827,70 | China 32968785,00 | China 100223,58 | China 0,00007 | China 8,18 |

Nations sorted by ratio of population per publications (A), H-Index (B) and citations (C). Additionally, nations sorted by ratio of publications per GDP (D) as well as citations per publications (E). Based on data in Table 1. Colors indicate the corresponding continent: Asia (yellow), Europe (blue), North America (red), Oceania (green); GDP: gross domestic product. USA: United States of America; UK: United Kingdom (query date: 20.10.19)

## Conclusion

This software-based approach allows, besides valid cleansing of metadata, to analyze scientometric topics of interests, e.g., nations, institutions, authors, and the possibility of performing gender analysis across different research areas. We used bioinformatic techniques and included contemporary visualization techniques such as heatmaps to illustrate global research performance and chord diagrams for cooperation between nations, institutes, and individual researchers. The simultaneous inclusion of data in tabular form also allows the results to be easily understood and made transparent, which is an essential part of our approach. This is illustrated by the analyses on percutaneous coronary interventions. Global research activities regarding percutaneous coronary intervention are dominated by the United States which is demonstrated by the highest number of publications, citations and H-Index (Table 1, Fig. 1a, b). If, however, the international research pattern is analyzed, Europe as a whole is leading, followed by North America and Asia (Fig. 1b, c). As our approach includes the underlying data provided in the tables section, research performance can be contextualized (Tables 1, 2). In this sense, relatively low populated countries such as the Netherlands, Switzerland, Denmark, Israel, and Sweden are, in this example, identified as particularly productive. Hence, it is interesting to notice that despite achieving a prominent position in terms of publication performance, the USA is not leading regarding relative aspects (Table 2). Interestingly, China, despite a high quantity of published articles, occupies only a low ranking in terms of quality. This demonstrates how research efforts can be visualized using the SciPE web application. Subsequently, further analyses are necessary to identify potential causes for such developments, e.g., productive research clusters and how to differentiate between high quantity versus high quality. As a result, scientometric analyses enrich our understanding of the global research architecture and may help to answer questions concerning research quantity and quality, such as who leads the world in science [25]. This may foster future research efforts and collaborations. Also, negative aspects such as excessive self-citations, authorships without sufficient contributions, and gender disparities may be identified and counteracted by the more frequent use of scientometric analyses [26]. As a result, the present study is an important step toward an improved quality of scientometric analyses with novel aspects such as plain visualization of international research networks/distribution, heatmapping based on exact geocoordinates, and central implementation of tables including specific ratios. The possibilities of applying this technique to different areas of research are only limited to the availability of underlying bibliographic data. Also, a major advantage of the graph-based structure is that it allows manual user-driven corrections without the need of a rule-based reparsing of the data to perform valid analyses. In summary, SciPE provides a user-friendly, transparent, and practicable approach to create high-quality scientometric research.

## Compliance with ethical standards

## References

1. Van Noorden R (2014) Global scientific output doubles every nine years. Nature news blog. http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html. Accessed 31 Aug 2019
2. Martín-Martín A, Orduna-Malea E, Thelwall M, Delgado López-Cózar E (2018) Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories. J Informetr 12:1160–1177. https://doi.org/10.1016/j.joi.2018.09.002
3. Clarivate Analytis > Web of Science group. The world´s largest publisher-neutral citation index and research intelligence platform. https://clarivate.com/webofsciencegroup/solutions/web-of-science/. Accessed 31 Aug 2019
4. Chen C, Hu Z, Liu S, Tseng H (2012) Emerging trends in regenerative medicine: a scientometric analysis in *CiteSpace*. Expert Opin Biol Ther 12:593–608. https://doi.org/10.1517/14712598.2012.674507
5. van Eck NJ, Waltman L (2014) CitNetExplorer: a new software tool for analyzing and visualizing citation networks. J Informetr 8:802–823. https://doi.org/10.1016/j.joi.2014.07.006
6. van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84:523–538. https://doi.org/10.1007/s11192-009-0146-3
7. Grauwin S, Jensen P (2011) Mapping scientific institutions. Scientometrics 89:943–954. https://doi.org/10.1007/s11192-011-0482-y
8. Groneberg-Kloft B, Fischer TC, Quarcoo D, Scutaru C (2009) New quality and quantity indices in science (NewQIS): the study protocol of an international project. J Occup Med Toxicol 4:16. https://doi.org/10.1186/1745-6673-4-16
9. Suresh S (2012) Global challenges need global solutions. Nature 490:18. https://www.nature.com/articles/490337a.pdf. Accessed 31 Aug 2019
10. Bendels MHK, Brüggmann D, Schöffel N, Groneberg DA (2016) Gendermetrics.NET: a novel software for analyzing the gender representation in scientific authoring. J Occup Med Toxicol. http://occup-med.biomedcentral.com/articles/10.1186/s12995-016-0133-6. Accessed 31 Aug 2019
11. Menzel LC, Kramer PW, Groneberg DA, Bendels MHK (2019) Gender disparities in authorships of Alzheimer's disease and dementia research articles. J Alzheimer's Dis. https://doi.org/10.3233/JAD-190216

12. Kayvanpour E, Sedaghat-Hamedani F, Gi W-T et al (2019) Clinical and genetic insights into non-compaction: a meta-analysis and systematic review on 7598 individuals. Clin Res Cardiol. https://doi.org/10.1007/s00392-019-01465-3

13. Wernly B, Seelmaier C, Leistner D et al (2019) Mechanical circulatory support with Impella versus intra-aortic balloon pump or medical treatment in cardiogenic shock—a critical appraisal of current data. Clin Res Cardiol. https://doi.org/10.1007/s00392-019-01458-2

14. Jobs A, Abdin A, de Waha-Thiele S et al (2019) Angiotensin-converting-enzyme inhibitors in hemodynamic congestion: a meta-analysis of early studies. Clin Res Cardiol. https://doi.org/10.1007/s00392-019-01456-4

15. Vargas KG, Haller PM, Jäger B et al (2019) Variations on classification of main types of myocardial infarction: a systematic review and outcome meta-analysis. Clin Res Cardiol 108:749–762. https://doi.org/10.1007/s00392-018-1403-3

16. Shah SR (2019) Interventional closure vs. medical therapy of patent foramen ovale for secondary prevention of stroke: updated meta-analysis. Clin Res Cardiol 108:452. https://doi.org/10.1007/s00392-018-1382-4

17. Lorenzi M, Ambegaonkar B, Baxter CA et al (2019) Ezetimibe in high-risk, previously treated statin patients: a systematic review and network meta-analysis of lipid efficacy. Clin Res Cardiol 108:487–509. https://doi.org/10.1007/s00392-018-1379-z

18. Scholz SS, Vukadinović D, Lauder L et al (2019) Effects of arteriovenous fistula on blood pressure in patients with end-stage renal disease: a systematic meta-analysis. J Am Heart Assoc. https://doi.org/10.1161/JAHA.118.011183

19. Murray M-I, Arnold A, Younis M et al (2018) Cryoballoon versus radiofrequency ablation for paroxysmal atrial fibrillation: a meta-analysis of randomized controlled trials. Clin Res Cardiol 107:658–669. https://doi.org/10.1007/s00392-018-1232-4

20. Sedaghat-Hamedani F, Kayvanpour E, Tugrul OF et al (2018) Clinical outcomes associated with sarcomere mutations in hypertrophic cardiomyopathy: a meta-analysis on 7675 individuals. Clin Res Cardiol 107:30–41. https://doi.org/10.1007/s00392-017-1155-5

21. US National Library of Medicine > National Institute of health > PubMed. https://www.ncbi.nlm.nih.gov/pubmed/. Accessed 31 Aug 2019

22. UniRank > Universities search engine. https://www.4icu.org. Accessed 31 Aug 2019

23. Python > SexMaschine 0.1.1. https://pypi.org/project/SexMachine/. Accessed 31 Aug 2019

24. Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA 102:16569–16572. https://doi.org/10.1073/pnas.0507655102

25. Mallapaty S (2018) Not so fast. Who really leads the world in science? Nature Index. https://www.natureindex.com/news-blog/not-so-fast-who-really-leads-the-world-in-science. Accessed 31 Aug 2019

26. Ioannidis PA, Klavans R, Boyack KW (2018) Thousands of scientists publish a paper every five days. https://www.nature.com/articles/d41586-018-06185-8. Accessed 31 Aug 2019

27. Central Intelligence Agency (2018) The world Factbook 2018. https://www.cia.gov/library/publications/download/download-2018/index.html