RESEARCH ARTICLE

Comprehensive mapping of local and diaspora scientists: A database and analysis of 63,951 Greek scientists

John P. A. Ioannidis[1], Chara Koutsioumpa[2], Angeliki Vakka[3], Georgios Agoranos[4], Chrysanthi Mantsiou[5], Maria Kyriaki Drekolia[6], Nikos Avramidis[7], Despina G. Contopoulos-Ioannidis[8], Konstantinos Drosatos[9,10], and Jeroen Baas[11]

[1]Stanford Prevention Research Center, Department of Medicine, and Department of Epidemiology and Population Health, Stanford University School of Medicine and Meta-Research Innovation Center at Stanford (METRICS), Stanford, California, USA
[2]Department of Neurobiology, Harvard Medical School, Boston, USA
[3]School of Medicine, University of Patras, Greece
[4]Hellenic Center of Mental Health and Research (EKEPSYE), Athens, Greece
[5]Clinical Research and Evidence-Based Medicine Unit, Second Medical Department, Aristotle University of Thessaloniki, Thessaloniki, Greece
[6]Institute for Vascular Signalling, Centre for Molecular Medicine, Goethe University, Frankfurt am Main, Germany
[7]Roslin Institute, University of Edinburgh
[8]Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA
[9]Department of Pharmacology, Center for Translational Medicine, Lewis Katz School of Medicine, Temple University, Philadelphia, PA, USA
[10]ARISTEiA-Institute for the Advancement of Research & Education in Arts, Sciences & Technology, McLean, VA, USA
[11]Elsevier, Amsterdam, Netherlands

an open access journal

Citation: Ioannidis, J. P. A., Koutsioumpa, C., Vakka, A., Agoranos, G., Mantsiou, C., Drekolia, M. K., Avramidis, N., Contopoulos-Ioannidis, D. G., Drosatos, K., & Baas, J. (2021). Comprehensive mapping of local and diaspora scientists: A database and analysis of 63,951 Greek scientists. *Quantitative Science Studies*, *2*(2), 733–752. https://doi.org/10.1162/qss_a_00136

DOI:
https://doi.org/10.1162/qss_a_00136

Peer Review:
https://publons.com/publon/10.1162/qss_a_00136

Received: 21 January 2021
Accepted: 24 March 2021

Corresponding Author:
John P. A. Ioannidis
jioannid@stanford.edu

Handling Editor:
Ludo Waltman

The MIT Press

Keywords: brain drain, diaspora, Greece, science mapping, scientific workforce

ABSTRACT

Research policy and planning for a given country may benefit from reliable data on both its scientific workforce as well as the diaspora of scientists for countries with a substantial brain drain. Here we use a systematic approach using Scopus to generate a comprehensive country-level database of all scientists in Greece. Moreover, we expand that database to include also Greek diaspora scientists. The database that we have compiled includes 63,951 scientists who have published at least five papers indexed in Scopus. Of those, 35,116 have an affiliation in Greece. We validate the sensitivity and specificity of the database against different control sets of scientists. We also analyze the scientific disciplines of these scientists according to the Science Metrix classification (174 subfield disciplines) and provide detailed data on each of the 63,951 scientists using multiple citation indicators and a composite thereof. These analyses demonstrate differential concentrations in specific subfields for the local versus the diaspora cohorts, as well as an advantage of the diaspora cohort in terms of citation indicators, especially among top-impact researchers. The approach that we have taken can also be applied to map the scientific workforce of other countries and nations for evaluation, planning, and policy purposes.

1. INTRODUCTION

The construction of scientist databases can be a useful tool for evaluation, planning, and policy-making related to science. Efforts to compile national databases of scientists with performance metrics, in particular citation indices, are sometimes undertaken by research assessment authorities (Moed, 2008; Rijke, Wouters et al., 2016). Often these efforts may not be sufficiently inclusive. For example, they may depend on nonsystematic efforts where scientists voluntarily contribute information themselves to be included. Moreover, citation metrics are difficult to standardize, especially when they are not calculated according to the same processes for all scientists, and when differences between scientific fields are insufficiently accounted for. Importantly, for many countries, brain drain is a major challenge for their scientific workforce (Doria Arrieta, Pammolli, & Petersen, 2017; Ioannidis, 2004; Veugelers, 2017). In these countries, planning and policy decisions would greatly benefit from mapping not only the disciplines and impact of scientists who still work in the country but also of those who have emigrated elsewhere. First-generation emigrating scientists and often even second and higher generation emigrants may often still be interested in engaging with the scientific workforce of their country of origin, thus contributing valuable expertise. Countries with strong diaspora may benefit from the skills of diaspora scientists. Scientific diaspora can be useful for both the mobile scientists (Halevi, Moed, & Bar-Ilan, 2016; Petersen, 2018; Robinson-Garcia, Sugimoto, et al., 2019) and the countries involved at both ends, as it can constitute a modern tool of scientific diplomacy and cooperation between the two countries (Stark, Helmenstein & Prskawetz, 1997; Wagner & Jonkers, 2017).

Here we demonstrate how a large-scale standardized approach can be used to create an inclusive, comprehensive database of scientists in a specific nation. Moreover, we show how one can expand that database to include also scientists who have migrated to other countries. We focus our efforts on Greece and its national workforce and scientific diaspora. Greece is a country that has sustained a very strong current of brain drain over the years (Ifanti, Argyriou et al., 2014; Moris, Karachaliou, & Kontos, 2017; Theodoropoulos, Kyridis et al., 2014; Trachana, 2013). Moreover, the country has been hit by a major economic crisis that has severely limited funding for research and development. Despite some improvements in recent years, funding remains highly suboptimal. Furthermore, scientists of Greek origin include many extremely influential scientists worldwide and past analyses suggest that there are many high-impact Greek scientists, both in Greece and abroad, who are leaders in their fields (Yurte, 2017). Moreover, such previous work has suggested that the number of Greek scientists with substantial impact is much higher proportionately than the share of Greeks in the global population (10 million in Greece and perhaps another 3 million in the diaspora) (Yurte, 2017). Mapping Greek scientists in Greece and worldwide would be a valuable resource. The availability of comprehensive science publications databases such as Scopus and the fact that many Greek first names and a large majority of Greek last names tend to be highly specific for Greek descent allow the creation of a database of scientists of Greek origin. In this paper, we describe how we have constructed such a database and how we have examined its sensitivity and specificity in validation samples. We also present descriptive data for the entire database and for comparative evaluations of Greek origin scientists who have an affiliation in Greece and for those who have an affiliation in other countries. Our work may offer a template for similar scientist-mapping efforts on other countries.

2. METHODS

2.1. Eligibility Criteria

We aimed to capture all scientists of Greek origin who have at least five published papers (articles, reviews, and conference proceedings). Eligible scientists were both those born in Greece and those born elsewhere (second or higher generation), but whose family had a Greek origin. Scientists were eligible regardless of whether they had their current main affiliation in Greece or elsewhere. We excluded scientists who had fewer than two papers published after 1950.

The overall strategy aimed at finding typical Greek names first and then retrieving all author profiles with these names. To capture eligible scientists with an affiliation in Greece, we queried Scopus (Baas, Schotten et al., 2020) as of January 15, 2020 and identified all the last names that had at least one author ID (with any number of papers assigned) that included an affiliation in Greece. We found 70,967 names where at least one author ID has an affiliation address in Greece. One researcher manually screened all of these names to identify those that seemed to be of Greek origin, allowing for inclusion of those who might be probable, to avoid losing potentially eligible names. A second researcher then examined the manual extraction and made amendments. Eventually, 57,732 last names were retained.

We also screened manually the files of the top 100,000 most cited scientists based on a composite indicator that had been published previously (Ioannidis, Baas et al., 2019). We used three different files of the top-cited scientists, each of which captured the top 100,000 including self-citations as well as the top 100,000 excluding self-citations based on career-long data in Scopus until the end of 2017 (https://dx.doi.org/10.17632/btchxktzyw.1#file-ad4249ac-f76f-4653-9e42-2dfebe5d9b01); based on citations received during a single calendar year (2017) (https://dx.doi.org/10.17632/btchxktzyw.1#file-b9b8c85e-6914-4b1d-815e-55daefb64f5e); and based on career-long data until the end of 2018 (https://dx.doi.org/10.17632/btchxktzyw.1#file-bade950e-3343-43e7-896b-fb2069ba3481). These three files manually yielded 1,044, 990, and 1,013 eligible authors of Greek origin, with large overlap between the three lists.

In addition, two online sources of common Greek first names (forebears.io and www.studentsoftheworld.info/penpals/stats.php?Pays=GRE) were screened manually starting from the most common ones until 86 first names were selected that were thought to be relatively specific for Greek origin people. For example, George is a common name in Greece, but it is not Greek specific (i.e., the vast majority of people with first name George are not of Greek origin). Conversely, Georgios is highly Greek specific.

At the next step, we retrieved from Scopus all author ID files with at least five papers (articles, reviews, or conference papers) where scientists had either a seemingly Greek-specific last name (any of the 57,332 last names mentioned above, or any of the last names of highly cited Greek scientists according to any of the three previously published lists) or a seemingly Greek-specific first name (any of the 86 mentioned above). Eventually, 124,656 author ID files were retrieved.

These 124,656 files were manually screened, perusing the information for each scientist, including the first name, last name, country of listed affiliation, and institution of listed affiliation that could help identify if the scientist was of Greek origin or not. The availability of all scientists who shared one of the seemingly Greek-specific names along with country information allowed us to identify whether any of these names were in fact not Greek-specific. Some last names occur identically both in Greeks and in some other nationality (e.g., Adam or Spinelli). In these cases, information on first name could help classify that individual if the first name was characteristically Greek. If the first name did not help to differentiate in this regard, the country information was used to arbitrate. The site https://forebears.io/ was consulted also in ambiguous cases, because it shows the relative frequency of surnames and names across different countries. If a scientist in a given country had a surname that appeared more frequently in Greece than in other countries, she or he was considered of Greek origin.

Of 124,656 author files, it was concluded that 62,837 were very likely Greek. We listed alphabetically by last names the 62,837 authors and recorded additional first names that seemed to be Greek specific. By screening 2,000 names at a time, it was found that relatively few new Greek-specific names were added after screening 8,000 authors and the incremental addition of eligible Greek origin authors would be limited by adding more first names. This process yielded 370 seemingly Greek-specific first names and we then searched Scopus for all additional author IDs with these first names that had not been already captured in the 62,837. These additional authors were then manually screened, and 1,012 were deemed (based on their name and country information) to be eligible. The resulting database, comprising 63,849 author IDs, was subjected to validation checks as described below. Additions and deletions emerging during these validation checks and a final contribution by the authors of the present study of Greek scientists they knew of but who had not been captured increased the final count by 102 to a final count of 63,951 author IDs.

The process is summarized in Figure 1.

2.2. Validation: Sensitivity for Capturing Scientists of Greek Origin Who Are in Greece

To evaluate the sensitivity of the compiled database in capturing scientists who work in Greece, we searched whether it had included scientists working at a university in Greece, the University of Thessaly. Scientists working in different universities and research institutions in Greece are not likely to have systematically different names, so one university is likely to provide a reasonably representative sample. We searched Google Scholar as the reference

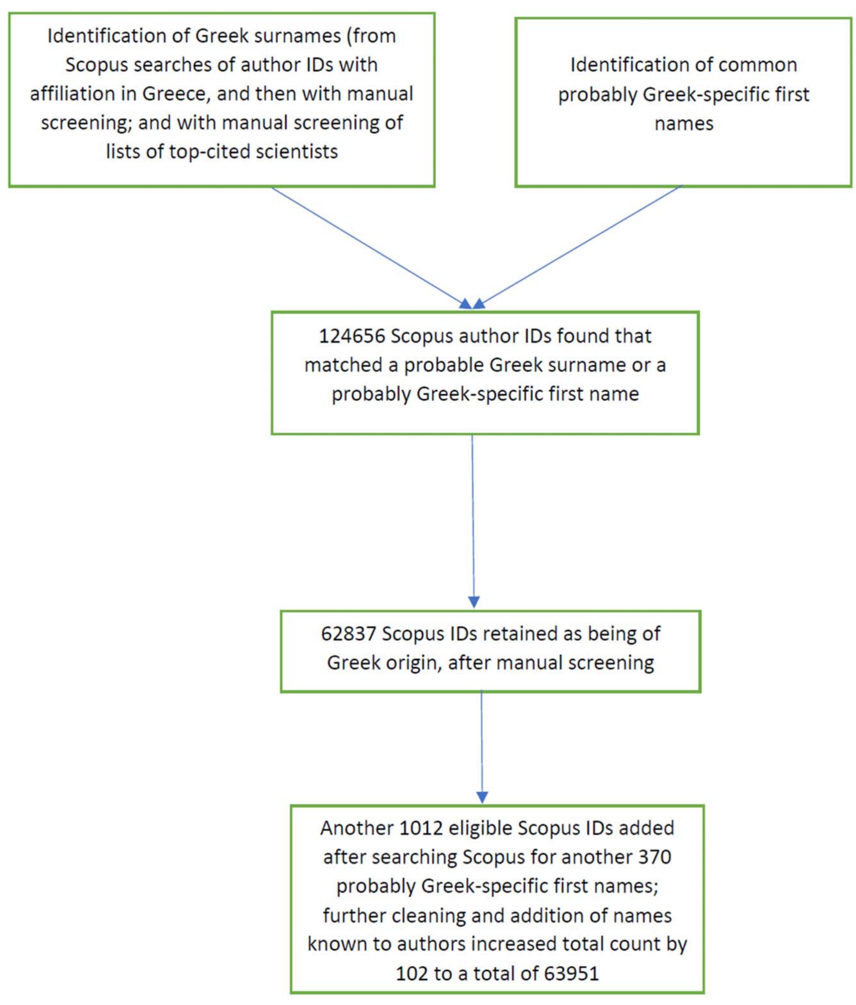

Figure 1. Flow diagram for identification of Greek scientists.

database, as scientists need to enter their names and affiliations by themselves in creating a profile in Google Scholar. The 130 most cited scientists with profiles and University of Thessaly affiliation in Google Scholar were screened and it was found that all of them (130/130) had been included in our compiled database. Therefore, the sensitivity was 100%, with binomial 95% confidence interval of 97.2% to 100%.

2.3. Validation: Sensitivity for Capturing Scientists of Greek Origin Who Are not in Greece

To evaluate the sensitivity of the compiled database in capturing scientists of Greek origin who do not work in Greece, we used two approaches.

First, we used a sample of scientists who had entered their names in a LinkedIn database of Greek biomedical scientists created by one of us (K. D.) for the World Hellenic Biomedical Association. We only considered names that had been entered by the scientists themselves, proving that they identified themselves as Greek; and we further limited the search to scientists who gave an address outside of Greece and who had a work title suggesting that they are faculty or other people in senior positions, as opposed to students. Of 42 such individuals, 34 were found to have at least five papers in Scopus. Of those 34, 26 were captured in the compiled database, for a sensitivity of 76.5% (95% confidence interval, 58.8% to 89.3%).

Second, we used the names of people listed in the Wikipedia entry on Greek Diaspora. These names are not necessarily of scientists; therefore we examined whether each of the names would have been captured either through one of the Greek-specific last names or through one of the first Greek-specific names that we had put together to compile our database of Greek scientists. For artists and other people who had acquired an artistic/stage name, we used their original name, as changing to artistic/stage names would not apply for scientists. We excluded from the screening people born before 1900, as Greek names in the remote past may have been different. Eventually, 28 first-generation and 88 second- or later-generation Greeks were eligible for screening. Of these 14/28 and 35/88 would have been captured by our last or first name searches, corresponding to sensitivities of 50% (95% confidence interval, 30.6% to 69.4%) and 39.8% (95% confidence interval, 29.4% to 50.8%), respectively.

The sensitivity estimates should be interpreted cautiously given the relatively small numbers and they leave some uncertainty about the total number of diaspora scientists.

2.4. Validation: Specificity for Capturing Scientists of Greek Origin

To evaluate whether the compiled database might have captured any scientists who were not actually Greek, we randomly selected 100 of the 63,849 author IDs. For each of them, we tried to find whether we could find their name written in Greek in the web. Of the 100, their Scopus affiliation was in Greece for 62, in Cyprus for four, and in other countries for 32; for two authors we had no listed affiliation in Scopus. We could find their name written in Greek for all 100 authors. Therefore, the specificity was 100% (95% confidence interval 96.4% to 100%).

2.5. Evaluation of Split Author Files

Some scientists in Scopus may have their published work split in two or more author ID files, and Scopus encourages authors to communicate directly with them to merge such split files. To assess how common this pattern might be in the compiled database of Greek authors, after listing the names alphabetically, every 600th name was selected and assessed whether more than one author ID files may exist for that person in the database. Of 106 screened names, nine (8.5%, 95% CI, 4.0 to 15.5%) had their work split into two files (n = 8) or three files (n = 1).

2.6. Data Included for Each Scientist in the Database

From each author ID file included in the database, the following information is included based on data directly imported from Scopus on October 1, 2020 (when 7,983,030 author ID files with at least five papers (articles, reviews, or conference papers) were available in Scopus) and calculations that are the same as those performed for a recently published list of top-cited scientists (Ioannidis, Boyack, & Baas, 2020): affiliation and country; publication year of earlier and latest Scopus-indexed publication; number of publications; number of publications in 1960–2020; six citation indicators (total citations, h-index, h_m-index, citations to single-authored publications, citations to first- or single-authored publications, citations to first-, single-, or last-authored publications) and their composite (all indicators being presented both with and without self-citations); proportion of self-citations; ratio of citations to citing papers; ranking according to the composite indicator among all scientists worldwide with at least five papers; most common field of publications according to the 22-field Science Metrix classification; two most common subfields of publications according to the 174-subfield Science Metrix classification; and ranking according to the composite indicator among all the scientists in the same main (most common) Science Metrix subfield. For details on the Science Metrix classification see Archambault, Beauchesne, and Caruso (2011) and Zhang, Zhao, and LeCun (2015). For authors where Scopus listed an affiliation but not a country, we tried to identify the country whenever it would be unambiguous based on the provided affiliation.

3. RESULTS

3.1. Main Descriptive Characteristics

Of the 63,951 author ID files included in the final database, country of affiliation was available for 63,174, and 35,116 (55.6%) of them had their affiliation in Greece. Large shares of this cohort

Figure 2. Worldwide distribution of scientists.

of scientists were also located in the United States (*n* = 9,339, 14.8%), United Kingdom (*n* = 6,165, 9.8%), Germany (*n* = 2,083, 3.3%), Cyprus (*n* = 1,688, 2.7%), Australia (*n* = 1,155, 1.8%), France (*n* = 1,141, 1.8%), Canada (*n* = 1,110, 1.8%), and Switzerland (*n* = 994, 1.6%), but the diaspora was worldwide (Figure 2).

A total of 12,299 (19.2%) scientists had published their first Scopus-indexed paper after 2010 and 38,248 (59.8%) had been recently active, publishing their last paper in 2018 or later. The median number of published papers was 13 (interquartile range 8 to 31) and the median number of citations was 153 (interquartile range 52 to 478).

As shown in Table 1, scientists with affiliation in Greece had a similar number of papers to scientists with affiliation outside of Greece, but they had substantially fewer citations and fewer papers that cited their work, and were placed on average in lower ranks compared with scientists with affiliation outside of Greece. The results were qualitatively similar regardless of whether self-citations were counted or excluded (Table 1). Scientists with affiliation outside of Greece tended to have younger publication ages (median for year of first publication 2004 versus 2002).

A total of 33,956 scientists with affiliation in Greece and 26,150 scientists with affiliation in other countries could be assigned to a main scientific subfield. Among scientists who were in the top 0.1% of their subfield, the vast majority (86%) of them had an affiliation outside of Greece rather than in Greece (96 versus 15). For the top 0.5%, the respective numbers were 348 versus 89, for the top 1% the respective numbers were 648 versus 250, and for the top 5% the respective numbers were 2,438 versus 1,724, always with a strong preponderance of

Table 1. Characteristics of scientists according to their country of current affiliation

Characteristic, median (IQR)	Affiliation in Greece *n* = 35,116	Affiliation in other country *n* = 28,058
Number of papers	13 (7–30)	14 (8–32)
Year of first paper	2002 (1993–2008)	2004 (1994–2010)
Year of most recent paper	2019 (2013–2020)	2019 (2014–2020)
Ranking across all science, thousands (excluding self-citations)	3,505 (1,691–5,463)	2,812 (1,265–4,730)
Citations (excluding self-citations)	139 (50–408)	184 (59–609)
Citing papers (excluding self-citations)	131 (48–375)	169 (55–541)
Citations to citing papers ratio (excluding self-citations)	1.04 (1.00–1.11)	1.07 (1.02–1.16)
Percentage of self-citations	14 (7–25)	15 (8–25)
Ranking across all science, thousands (with self-citations)	3,508 (1,681–5,458)	2,801 (1,259–4,697)
Citations (with self-citations)	169 (63–486)	226 (76–730)
Ranking in main subfield (with self-citations)	28,711 (11,338–61,183)	21,378 (7,500–48,629)
Ranking in main subfield (without self-citations)	28,693 (11,312–61,180)	21,324 (7,535–48,751)
Percentile in main subfield (with self-citations)	46 (22–71)	37 (16–61)
Percentile in main subfield (without self-citations)	46 (22–71)	37 (16–61)

Table 2. Number of scientists in each scientific subfield

Scientific subfield	Country			
	Greece	Other	Unknown	Total
Accounting	20	32	0	52
Acoustics	112	154	0	266
Aerospace & Aeronautics	57	159	3	219
Agricultural Economics & Policy	30	27	0	57
Agronomy & Agriculture	211	54	0	265
Allergy	80	62	0	142
Analytical Chemistry	368	194	2	564
Anatomy & Morphology	35	17	0	52
Anesthesiology	130	81	0	211
Anthropology	16	40	0	56
Applied Ethics	7	18	1	26
Applied Mathematics	160	100	1	261
Applied Physics	632	586	0	1,218
Archaeology	107	61	0	168
Architecture	5	7	0	12
Art Practice, History & Theory	1	6	0	7
Arthritis & Rheumatology	194	126	0	320
Artificial Intelligence & Image Processing	1,685	1,293	7	2,985
Astronomy & Astrophysics	166	246	1	413
Automobile Design & Engineering	1	5	1	7
Behavioral Science & Comparative Psychology	6	20	0	26
Biochemistry & Molecular Biology	286	474	0	760
Bioinformatics	72	105	0	177
Biomedical Engineering	262	248	1	511
Biophysics	64	71	0	135
Biotechnology	181	128	0	309
Building & Construction	168	148	0	316
Business & Management	200	203	0	403
Cardiovascular System & Hematology	1,671	768	10	2,449
Chemical Engineering	220	236	0	456
Chemical Physics	200	234	1	435

Table 2. (*continued*)

Scientific subfield	Country			
	Greece	Other	Unknown	Total
Civil Engineering	249	251	2	502
Classics	39	49	1	89
Clinical Psychology	24	51	0	75
Communication & Media Studies	9	56	0	65
Complementary & Alternative Medicine	6	8	0	14
Computation Theory & Mathematics	123	119	1	243
Computer Hardware & Architecture	173	164	0	337
Criminology	7	45	0	52
Cultural Studies	0	11	0	11
Dairy & Animal Science	214	54	0	268
Demography	13	10	0	23
Dentistry	421	287	5	713
Dermatology & Venereal Diseases	209	150	1	360
Design Practice & Management	24	35	0	59
Development Studies	3	13	0	16
Developmental & Child Psychology	35	94	0	129
Developmental Biology	172	712	0	884
Distributed Computing	64	94	0	158
Drama & Theater	3	2	0	5
Ecology	119	72	1	192
Econometrics	8	13	0	21
Economic Theory	4	22	0	26
Economics	310	286	0	596
Education	456	329	0	785
Electrical & Electronic Engineering	293	267	0	560
Emergency & Critical Care Medicine	183	93	0	276
Endocrinology & Metabolism	577	342	9	928
Energy	869	591	7	1,467
Entomology	195	48	0	243
Environmental & Occupational Health	14	25	0	39
Environmental Engineering	246	119	1	366

Table 2. (*continued*)

Scientific subfield	Country			
	Greece	Other	Unknown	Total
Environmental Sciences	534	163	0	697
Epidemiology	43	30	0	73
Evolutionary Biology	64	57	0	121
Experimental Psychology	39	138	1	178
Family Studies	4	3	0	7
Finance	95	159	0	254
Fisheries	146	47	0	193
Fluids & Plasmas	159	195	0	354
Folklore	2	0	0	2
Food Science	377	132	4	513
Forestry	80	44	0	124
Gastroenterology & Hepatology	661	252	2	915
Gender Studies	5	7	0	12
General & Internal Medicine	375	231	5	611
General Chemistry	12	51	0	63
General Clinical Medicine	102	38	0	140
General Mathematics	237	162	0	399
General Physics	131	151	0	282
General Psychology & Cognitive Sciences	2	5	0	7
Genetics & Heredity	96	174	0	270
Geochemistry & Geophysics	270	167	8	445
Geography	37	26	0	63
Geological & Geomatics Engineering	272	187	1	460
Geology	27	20	0	47
Geriatrics	26	32	0	58
Gerontology	17	17	0	34
Health Policy & Services	46	78	0	124
History	19	39	1	59
History of Science, Technology & Medicine	18	4	0	22
History of Social Sciences	2	7	0	9
Horticulture	35	6	0	41

Table 2. (*continued*)

Scientific subfield	Country			
	Greece	Other	Unknown	Total
Human Factors	28	77	0	105
Immunology	706	487	1	1,194
Industrial Engineering & Automation	332	390	0	722
Industrial Relations	10	14	0	24
Information & Library Sciences	38	34	1	73
Information Systems	143	204	4	351
Inorganic & Nuclear Chemistry	250	128	1	379
International Relations	16	49	0	65
Languages & Linguistics	57	57	0	114
Law	12	82	2	96
Legal & Forensic Medicine	30	23	0	53
Literary Studies	12	27	0	39
Logistics & Transportation	173	155	1	329
Marine Biology & Hydrobiology	247	83	1	331
Marketing	56	75	0	131
Materials	390	280	1	671
Mathematical Physics	38	26	0	64
Mechanical Engineering & Transports	253	218	1	472
Medical Informatics	184	78	0	262
Medicinal & Biomolecular Chemistry	236	149	0	385
Meteorology & Atmospheric Sciences	306	223	0	529
Microbiology	896	367	4	1,267
Microscopy	6	4	0	10
Mining & Metallurgy	50	28	2	80
Music	6	13	0	19
Mycology & Parasitology	56	38	1	95
Nanoscience & Nanotechnology	75	134	0	209
Networking & Telecommunications	1,100	953	10	2,063
Neurology & Neurosurgery	742	962	3	1,707
Nuclear & Particle Physics	509	623	0	1,132
Nuclear Medicine & Medical Imaging	575	377	1	953

Table 2. (*continued*)

Scientific subfield	Country			
	Greece	Other	Unknown	Total
Numerical & Computational Mathematics	92	51	0	143
Nursing	147	89	1	237
Nutrition & Dietetics	230	126	0	356
Obstetrics & Reproductive Medicine	636	303	1	940
Oceanography	102	43	0	145
Oncology & Carcinogenesis	1,718	868	3	2,589
Operations Research	138	118	0	256
Ophthalmology & Optometry	301	308	0	609
Optics	74	155	0	229
Optoelectronics & Photonics	258	319	1	578
Organic Chemistry	232	212	0	444
Ornithology	17	4	0	21
Orthopedics	461	348	1	810
Otorhinolaryngology	206	128	0	334
Paleontology	54	34	0	88
Pathology	123	70	1	194
Pediatrics	226	113	2	341
Pharmacology & Pharmacy	396	208	0	604
Philosophy	12	44	2	58
Physical Chemistry	172	86	4	262
Physiology	26	59	0	85
Plant Biology & Botany	498	161	0	659
Political Science & Public Administration	55	107	0	162
Polymers	282	222	1	505
Psychiatry	277	265	2	544
Psychoanalysis	7	21	3	31
Public Health	87	139	0	226
Rehabilitation	36	68	0	104
Religions & Theology	7	17	0	24
Respiratory System	534	255	1	790
Science Studies	22	29	0	51

Table 2. (*continued*)

Scientific subfield	Country			
	Greece	Other	Unknown	Total
Social Psychology	30	74	1	105
Social Sciences Methods	4	12	0	16
Social Work	5	21	0	26
Sociology	12	18	0	30
Software Engineering	66	126	0	192
Speech-Language Pathology & Audiology	37	66	0	103
Sport Sciences	331	104	0	435
Sport, Leisure & Tourism	44	75	0	119
Statistics & Probability	144	132	0	276
Strategic, Defence & Security Studies	160	98	1	259
Substance Abuse	18	36	0	54
Surgery	760	493	16	1,269
Toxicology	114	81	1	196
Tropical Medicine	34	32	0	66
Urban & Regional Planning	66	51	0	117
Urology & Nephrology	549	223	1	773
Veterinary Sciences	164	71	2	237
Virology	85	154	1	240
Zoology	33	18	0	51

scientists who were not in Greece (79%, 72%, and 59%, respectively, for these three thresholds). Below the top 5%, there was more equilibrium between scientists with affiliation outside of Greece versus in Greece, with the respective numbers being 7,842 versus 7,807 for the top 20%.

3.2. Scientific Fields

As shown in Table 2, Greek scientists had different representation across the 174 main scientific subfields of the Science Metrix classification. A number of fields of clinical medicine, biology, and agriculture/fisheries/forestry are more heavily represented for scientists who are in Greece, while the diaspora is more prominently represented in several social and economic sciences and some cutting-edge biomedical fields. For 25 subfields, scientists in Greece exceeded by more than 2:1 scientists with affiliation outside of Greece (Anatomy & Morphology, Environmental Engineering, Respiratory System, Obstetrics & Reproductive Medicine, Cardiovascular System & Hematology, Veterinary Sciences, Medical Informatics, Oceanography, Microbiology, Urology & Nephrology, Gastroenterology & Hepatology, General Clinical Medicine, Food Science, Marine Biology & Hydrobiology, Plant Biology &

Table 3. Scientists who are among the top 15 of their scientific subfield according to a composite citation indicator, excluding self-citations

Name of scientist	Affiliation	Country	Subfield	Rank*	*n* in subfield**	First degree
Santamouris, Mattheos	University of New South Wales (UNSW) Australia	aus	Building & Construction	1	27299	U Patras
Peppas, Nicholas A.	The University of Texas at Austin	usa	Pharmacology & Pharmacy	1	95625	NTUA
Terzopoulos, Demetri	University of California, Los Angeles	usa	Software Engineering	1	21515	McGill
Nicolaides, Kypros H.	King's College Hospital	gbr	Obstetrics & Reproductive Medicine	2	66792	King's College
Papadimitriou, Christos H.	Columbia University in the City of New York	usa	Computation Theory & Mathematics	2	16762	NTUA
Ioannidis, John P. A.	Stanford University School of Medicine	usa	General & Internal Medicine	2	107698	U Athens
Stamatakis, Alexandros	Karlsruhe Institute of Technology	deu	Bioinformatics	3	18635	TU Munich
Joannopoulos, John	Massachusetts Institute of Technology	usa	Optoelectronics & Photonics	3	102335	UC Berkeley
Alivisatos, A. Paul	University of California, Berkeley	usa	Nanoscience & Nanotechnology	4	75646	U Chicago
Ntziachristos, Vasilis	Helmholtz Center Munich German Research Center for Environmental Health	deu	Nuclear Medicine & Medical Imaging	5	84992	Aristotle U
Guibas, Leonidas J.	Stanford University	usa	Software Engineering	5	21515	CalTech
Buhalis, Dimitrios	Bournemouth University	gbr	Sport, Leisure & Tourism	6	6295	U Aegean
Giannelis, Emmanuel	Cornell University	usa	Polymers	6	81179	U Athens
Kanatzidis, Mercouri G.	Northwestern University	usa	Materials	6	180221	Aristotle U
Simopoulos, Artemis P.	Center for Nutrition, Genetics & Health	usa	Nutrition & Dietetics	8	35890	Barnard College
Bertsekas, Dimitri	Arizona State University	usa	Operations Research	9	23674	NTUA
Pavlou, Paul	C. T. Bauer College of Business	usa	Information Systems	9	16831	Rice U
Stephanopoulos, Gregory	Massachusetts Institute of Technology	usa	Biotechnology	9	50679	NTUA
Nicolaou, K. C.	Rice University	usa	Organic Chemistry	9	112004	U London
Diamantopoulos, Adamantios	Universitat Wien	aut	Marketing	10	10516	Heriot-Watt U
Gazetas, G.	National Technical University of Athens	grc	Strategic, Defence & Security Studies	10	17396	NTUA

Antonarakis, Stylianos E.	Université de Genève Faculté de Médecine	che	Genetics & Heredity	11	32809	U Athens
Pratsinis, Sotiris E.	ETH Zürich	che	Chemical Engineering	12	56292	Aristotle U
Chrousos, George P.	National and Kapodistrian University of Athens	grc	Endocrinology & Metabolism	12	69452	U Athens
Kalogirou, Soteris A.	Cyprus University of Technology	cyp	Energy	13	188556	Higher Tech Inst
Yannakakis, Mihalis	Columbia University in the City of New York	usa	Computation Theory & Mathematics	13	16762	NTUA
Avouris, Phaedon	IBM Thomas J. Watson Research Center	usa	Applied Physics	13	226884	Aristotle U
Lyketsos, Constantine G.	Johns Hopkins Bayview Medical Center	usa	Geriatrics	13	9246	Northwestern
Argyropoulos, Dimitris S.	NC State University	usa	Forestry	14	24339	U London
Giannakis, Georgios B.	University of Minnesota Twin Cities	usa	Networking & Telecommunications	14	162693	NTUA
Davatzikos, Christos	University of Pennsylvania	usa	Nuclear Medicine & Medical Imaging	14	84992	NTUA
Karniadakis, George Em	Brown University	usa	Applied Mathematics	15	16040	MIT

* Rank among all scientists in the same subfield, regardless of whether they are alive or deceased. For example, in General & Internal Medicine the top ranked scientist is Douglas Altman, who is deceased. Also of note, the top 32 scientists who are highly ranked based on the percentile in their subfield (not shown here) include 23 of the 32 who are top-ranked based on the absolute threshold (top 15 in the subfield).

** Number of scientists in the same subfield, including both those who are alive and those who are deceased; it is not straightforward to identify how many are deceased. The count includes those who have at least five papers (articles, reviews, or conference proceedings) indexed in Scopus and who have some papers classified in one of the 174 Science Metrix subfields.

Botany, Fisheries, Sport Sciences, Environmental Sciences, Agronomy & Agriculture, Dairy & Animal Science, Entomology, Ornithology, History of Science, Technology & Medicine, Horticulture, Folklore). Conversely, in 32 subfields, scientists outside of Greece exceeded by more than 2: 1 scientists with affiliation in Greece (Cultural Studies, Law, Criminology, Communication & Media Studies, Art Practice, History & Theory, Economic Theory, Automobile Design & Engineering, Development Studies, General Chemistry, Social Work, Developmental Biology, Philosophy, Experimental Psychology, History of Social Sciences, Behavioral, Science & Comparative Psychology, International Relations, Psychoanalysis, Social Sciences Methods, Aerospace & Aeronautics, Human Factors, Developmental & Child Psychology, Applied Ethics, Anthropology, General Psychology & Cognitive Sciences, Social Psychology, Religions & Theology, Physiology, Literary Studies, Music, Clinical Psychology, Optics, History).

3.3. Top-Cited Greek Scientists Across Different Fields

Thirty-two Greek scientists were among the top 15 of their scientific subfield based on a citation indicator excluding self-citations (Table 3). Almost all of them (30/32, 94%) were listed by Scopus with an affiliation outside of Greece. Of the 32 scientists, information on place of birth could be found on 28 (except for Terzopoulos, Stamatakis, Pavlou, and Argyropoulos); three were born in Cyprus (Nicolaides, Nicolaou, Kalogirou), three were born in the United States (Ioannidis, Joannopoulos, Alivisatos), one was born in the United Kingdom (Lyketsos), and the remaining 21 had been born in Greece. Of the 32, 18 had received their first degree from an institution in Greece.

4. DISCUSSION

We have created and validated a database of scientists of Greek origin that may be helpful for evaluation, planning, and research policy purposes. It may also serve as a template for similar efforts to be undertaken for other countries to map their scientific workforce. The iterative approach that we followed may also have special added value for countries that have sustained heavy brain drain and/or that have a substantial scientific diaspora. The sensitivity and specificity achieved from such an approach in constructing scientist databases from different countries may vary depending on how unique first and last names are to geographic origin.

Our approach has tried to identify scientists originating from Greece regardless of their present or past affiliations. We have also probed the sensitivity and specificity of the database membership. The database is dependent on Scopus coverage, so scientists in fields not well covered by Scopus may be particularly underrepresented. The database includes close to 64,000 author ID files representing scientists who have published at least five papers. Given that some scientists have their publications split in more than one file, the database probably includes close to 60,000 unique scientists. Validation exercises suggest that it probably misses very few scientists who meet the productivity eligibility criteria and who have an affiliation in a Greek institution. Conversely, a more substantial proportion has been missed among those who have an affiliation in an institution outside of Greece. The estimate of the missingness in this regard varies according to different validation sets that we used. Based on scientists working abroad who on their own initiative offered to be included in a database of Greek scientists, about one in four scientists were missed with our approach. The percentage of missingness was higher based on a Wikipedia list of diasporeans, and even more when extending beyond first-generation emigrants. It is unavoidable that our approach would miss Greeks who acquire non-Greek names (upon second and subsequent generations) and for people who change their names (e.g., through marriage, by making the name less foreign-sounding in their new country, or other reasons) and those who have a Greek last name that was not among the ones we searched for. Some of these individuals may still be captured if they possess a highly Greek-specific first name among the list of first names that we screened for. Therefore, even though scientists with an affiliation in Greece were a slight majority in the compiled database, Greek origin scientists with an affiliation outside of Greece would probably be the majority if all Greek origin scientists could have been retrieved. The total of Greek origin scientists meeting the productivity eligibility criteria may be in the range of 80,000–100,000 (~1.00–1.25% of the global total). Conversely, a few scientists are included in the database by failed disambiguation. The validation process suggests that this situation is probably very uncommon.

The database reflects the large extent of general emigration from Greece as well as the massive brain drain that the country has sustained over the years, with accelerated rates in the last decade in conjunction with the economic crisis that hit Greece worse than almost any other highly developed country. We noted that the cohort of scientists with affiliation outside of Greece had on average younger publication ages, as revealed by the year of their first paper; half of them published their first paper in 2004 or more recently.

While citation indicators are quite high for the entire database averages, scientists with an affiliation outside of Greece have substantially stronger citation indicators and higher rankings in their fields compared with scientists with affiliation in Greece. The difference is more prominent among top-cited scientists, where 86% of the Greek origin scientists who are in the top 0.1% of their subfield are not in Greece. Similarly, almost all (94%) of the Greek origin scientists who are among the top 15 of their subfield are not in Greece. Of interest, is that the large majority of these extremely highly cited scientists were born in Greece, and the majority also received their first degree in Greece. This further demonstrates the power of the brain drain process. At the same time, scientists who have remained in Greece still include large numbers placed in the top 20% of their subfield. Thus, the local scientific workforce still has considerable capacity for excellence. Of note, given our search strategy, is that our database has practically 100% sensitivity for Greek scientists abroad who are in the top 2% of citation impact, while several scientists with lesser impact may have been missed.

The database includes scientists scattered across almost every scientific subfield. Scientists with an affiliation in Greece have stronger concentrations than those with affiliations outside of Greece in many fields of clinical medicine, several fields of biology, and agriculture/fisheries/forestry. Greece has one of the highest rates of physicians per population in the world, if not the highest (country-level data on medical doctors per 10,000 population are available at https://www.who.int/data/gho/data/indicators/indicator-details/GHO/medical-doctors-per-10-000-population). Many of them are engaged in research, authoring or coauthoring papers, as scientific publications are requested and appraised not only for academic track positions, but even for regular clinical positions in the national health system. The advantage is that these incentives create a large pool of physicians with exposure to research. The disadvantage is that much of this research may not be of high quality and these authors have no lasting commitment to research. The concentration in subfields of agriculture, fisheries, and biology is probably explained by the nature of the economy, although agriculture and related fields have shrunk in latest years. Conversely, there are several other fields where most scientists of Greek origin do not work in Greece. This pattern is particularly strong in the social and economic sciences and some cutting-edge biomedical sciences, such as developmental biology.

Some limitations need to be discussed. First, as we have already acknowledged, the database is still missing several Greek origin scientists, in particular among those living and working abroad. We encourage people to provide relevant information at www.drosatos.com/greekscientists to bring such cases to our attention. While it is impossible to update the database by adding one more scientist at a time, collecting information on missing individuals may allow us to consider further optimized automated processes in the future. While this paper was in peer review, and as of March 20, 2021, 13 additional names of Greek scientists were provided to us, but seven of those were already included in the database, two had fewer than five papers by early 2020, and only four had been missed (Anna-Bettina Haidich [Aristotle U Thessaloniki, $h = 29$], Elias Franses [Purdue, $h = 36$], Iosif Koutagiar [Hygheia Hospital, $h = 9$], and Christos Chinopoulos [Semmelweis Egyetem, $h = 31$]).

Second, the constructed database was restricted to scientists with at least five full papers. In the entire Scopus database, roughly four-fifths of author ID files have fewer than five papers. Some of the author ID files with sparse papers may be split-off fragments of the publication corpus of authors represented by some larger file. Nevertheless, by extrapolation, the total number of Greek authors who have published at least one paper may be in the range of 250,000–500,000. The overwhelming majority of authors of 1–4 papers are not major contributors or leaders in the scientific enterprise. However, many young scientists in this group may become major contributors or leaders in the future. Therefore, follow-up updates would be useful to perform.

Third, errors (either splitting the same author into two or more author ID files or including some papers by two or more authors in the same file) and inaccuracies in affiliations are possible. Authors who recognize errors should contact Scopus directly to make these corrections in Scopus itself, so that they may be carried over in our database with any potential future updates. The entire Scopus database currently has overall 98.1% precision (proportion of papers in an author ID file that belong to the author) and 94.4% recall (proportion of papers of an author included in the largest profile) (Baas et al., 2020). Precision and recall may be even better for Greek-name authors, because Greek names are more rare and thus more specific than those of most other origins (e.g., the disambiguation challenges for "Liu Wang" are greater than for "Yiannis Triantafyllou"). We found 8.5% of the authors in the database to have a split profile and, given that even when one profile carries the large majority of the author's papers, recall probably substantially exceeds 94.4% for our database.

Fourth, allocation of fields and subfields follows a well-established classification, but some scientists may have an almost equal number of papers in two or more fields, and the most common one may not fully capture their expertise. Their ranking would have been different had they been classified in a different subfield. Moreover, even within the same subfield, there are granular subsections with different citation densities.

Fifth, allocation of affiliation and country is performed automatically by Scopus picking just one affiliation from the most recent papers of each author. Some authors have multiple current affiliations, and some may have changed their affiliation recently. Again, we encourage authors who want to change their listed affiliation to communicate directly with Scopus. Misclassification may affect some authors in their classification as being in Greece versus outside of Greece. However, it would have been extremely difficult to curate affiliations manually and it is impossible for an outsider to know which of many affiliations an author may prefer.

Sixth, our database does not automatically distinguish between first, second, and higher generation emigrants. If deemed desirable, this would have to be done manually, and it may have implications for policy (not losing scientists versus attracting scientists). Second and higher generation emigrants are not necessarily a sign of brain drain, as they did not emigrate themselves.

Finally, all citation metrics have limitations and they should be used with caution and not as absolute indicators (Hicks, Wouters et al., 2015; Waltman, 2016; Wang, Veugelers, & Stephan, 2017). We made no effort to assess the quality of the published works. Some authors may rank high, but may have other reasons for concern (e.g., retracted papers, or implausibly high self-citation metrics or evidence for citation farms). These need to be carefully scrutinized on a case-by-case basis.

Acknowledging these caveats, the compiled database offers a tool that may be useful for both research and policy purposes. For a country that is trying to recover from a lengthy economic crisis and a superimposed crisis from the recent COVID-19 pandemic, realization of its scientific potential, deceleration and reversal of the brain drain and informed decision-making in the interface of science and society may offer substantial added value. The brain drain and diaspora do not need to have negative consequences for the home country; mapping of the scientific workforce and diaspora may help to maximize positive impact (Davenport, 2004; Wagner & Jonkers, 2017).

We also hope that the iterative approach used here may be applied also to map the scientific workforce and scientific diaspora of other countries/nations as well. Scopus data can readily identify scientists with affiliation in a given country. In the case of Greece, where few scientists immigrate to from other countries, almost all scientists with affiliation in Greece have Greek names. This would not be true for countries that attract many scientists from other countries, but usually it is more important to map the entire scientific workforce rather than just native scientists. The ability to map the diaspora of different countries depends on whether there are many first and last names that are country-specific. Specificity may vary substantially across countries and careful validation and cross-checking procedures should be applied accordingly.

AUTHOR CONTRIBUTIONS

John P. A. Ioannidis: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing—Original draft, Writing—Review & editing. Chara Koutsioumpa: Data curation, Investigation, Validation, Writing—Review & editing. Angeliki Vakka: Data curation, Investigation, Validation, Writing—Review & editing. Georgios Agoranos: Data curation, Investigation, Validation, Writing—Review & editing. Chrysanthi Mantsiou: Data curation, Investigation, Validation, Writing—Review & editing. Maria Kyriaki Drekolia: Data curation, Investigation, Validation, Writing—Review & editing. Nikos Avramidis: Data curation, Investigation, Validation, Writing—Review & editing. Despina G. Contopoulos-Ioannidis: Data curation, Formal analysis, Investigation, Validation, Visualization, Writing—Review & editing. Konstantinos Drosatos: Conceptualization, Data curation, Investigation, Project administration, Supervision, Validation, Writing—Review & editing. Jeroen Baas: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing—Review & editing.

COMPETING INTERESTS

Jeroen Baas is an employee of Elsevier, which runs Scopus.

FUNDING INFORMATION

No funding has been received for this research.

DATA AVAILABILITY

All the data on the 63,951 scientists are available in Mendeley at https://doi.org/10.17632/zbyctscmbn.1.

REFERENCES

Archambault, É., Beauchesne, O. H., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics (ISSI)*, pp. 66–77. Durban, South Africa.

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*, 377–386. https://doi.org/10.1162/qss_a_00019

Davenport, S. (2004). Panic and panacea: Brain drain and science and technology human capital policy. *Research Policy*, *33*, 617–630. https://doi.org/10.1016/j.respol.2004.01.006

Doria Arrieta, O. A., Pammolli, F., & Petersen, A. M. (2017). Quantifying the negative impact of brain drain on the integration of European science. *Science Advances*, *3*, e1602232. https://doi.org/10.1126/sciadv.1602232, PubMed: 28439544

Halevi, G., Moed, H. F., & Bar-Ilan, J. (2016). Researchers' mobility, productivity and impact: Case of top producing authors in seven disciplines. *Publishing Research Quarterly*, *32*, 22–37. https://doi.org/10.1007/s12109-015-9437-0

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*, 429–431. https://doi.org/10.1038/520429a, PubMed: 25903611

Ifanti, A. A., Argyriou, A. A., Kalofonou, F. H., & Kalofonos, H. P. (2014). Physicians' brain drain in Greece: A perspective on the reasons why and how to address it. *Health Policy*, *117*, 210–215. https://doi.org/10.1016/j.healthpol.2014.03.014, PubMed: 24794987

Ioannidis, J. P. (2004). Global estimates of high-level brain drain and deficit. *FASEB Journal*, *18*, 936–939. https://doi.org/10.1096/fj.03-1394lfe, PubMed: 15173104

Ioannidis, J. P., Baas, J., Klavans, R., & Boyack, K. W. (2019). A standardized citation metrics author database annotated for scientific field. *PLoS Biology*, *17*, e3000384. https://doi.org/10.1371/journal.pbio.3000384, PubMed: 31404057

Ioannidis, J. P. A., Boyack, K. W., & Baas, J. (2020). Updated science-wide author databases of standardized citation indicators. *PLOS Biology*, *18*, e3000918. https://doi.org/10.1371/journal.pbio.3000918, PubMed: 33064726

Moed, H. F. (2008). UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics*, *74*, 153–161. https://doi.org/10.1007/s11192-008-0108-1

Moris, D., Karachaliou, G. S., & Kontos, M. (2017) Residency training in Greece: Job dissatisfaction paves the way to brain drain. *Annals of Translational Medicine*, *5*, 123. https://doi.org/10.21037/atm.2017.03.03, PubMed: 28361088

Petersen, A. M. (2018). Multiscale impact of researcher mobility. *Journal of The Royal Society Interface*, *15*(146), 20180580. https://doi.org/10.1098/rsif.2018.0580, PubMed: 30257927

Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—A literature review. *Research Evaluation*, *25*, 161–169. https://doi.org/10.1093/reseval/rvv038

Robinson-Garcia, N., Sugimoto, C. R., Murray, D., Yegros-Yegros, A., Larivière, V., & Costas, R. (2019). The many faces of mobility: Using bibliometric data to measure the movement of scientists. *Journal of Informetrics*, *13*, 50–63. https://doi.org/10.1016/j.joi.2018.11.002

Stark, O., Helmenstein, C., & Prskawetz, A. (1997). A brain gain with a brain drain. *Economics Letters*, *55*, 227–234. https://doi.org/10.1016/S0165-1765(97)00085-2

Theodoropoulos, D., Kyridis, A., Zagkos, C., & Konstantinidou, Z. (2014). "Brain Drain" phenomenon in Greece: Young Greek scientists on their way to immigration, in an era of "crisis." Attitudes, opinions and beliefs towards the prospect of migration. *Journal of Education and Human Development*, *3*, 229–248. https://doi.org/10.15640/jehd.v3n4a21

Trachana, V. (2013). Austerity-led brain drain is killing Greek science. *Nature*, *496*, 271. https://doi.org/10.1038/496271a, PubMed: 23598305

Veugelers, R. (2017). Countering European brain drain. *Science*, *356*(6339), 695–696. https://doi.org/10.1126/science.aan3920, PubMed: 28522487

Wagner, C., & Jonkers, K. (2017). Open countries have strong science. *Nature*, *550*, 32–33. https://doi.org/10.1038/550032a, PubMed: 28980660

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*, 365–391. https://doi.org/10.1016/j.joi.2016.02.007

Wang, J., Veugelers, R., Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*, 1416–1436. https://doi.org/10.1016/j.respol.2017.06.006

Yurte, T. (2017). An analysis of the foreign-educated elite academics in the United States. *Journal of Informetrics*, *11*, 358–370. https://doi.org/10.1016/j.joi.2017.02.008

Zhang, X., Zhao, J., & LeCun, Y. (2015) Character-level convolutional networks for text classification. In *Proceedings of the 28th International Conference on Neural Information Processing Systems* (Vol. 1, pp. 649–657).