TRAMES, 2019, **23(73/68)**, 3, 287–308

RESEARCH CONSORTIA DETERMINE A SIGNIFICANT PART OF THE BIBLIOMETRIC VISIBILITY OF ESTONIAN SCIENCE

Tanel Hirv

University of Tartu

Abstract. The study looks from a bibliometric perspective on how international collaboration affects Estonian science. The metrics for the study are retrieved using online InCites database (Clarivate Analytics). Results show that research consortia determine a significant part of the bibliometric visibility. An increase in impact from collaborations in consortia was largest in molecular biology and genetics, clinical medicine, and physics. Unfortunately, collaborations in consortia have also created a challenge from an evaluation viewpoint on how to measure scientific impact. Are hyper-authored articles worth as much as ordinary articles? If collaborations in consortia were excluded, then the largest increase in impact from internationalization over the observed period was seen in space science, followed by immunology, psychiatry/psychology, pharmacology, and toxicology.

Keywords: international collaboration, scientometrics, scientific impact, Estonian publications, hyperauthorship

DOI: https://doi.org/10.3176/tr.2019.3.02

1. Introduction

Over the last decades, scientific collaboration has become a major topic in science policy. It is widely believed that collaboration is ‘a good thing’ and it should be encouraged in every way possible (Katz and Martin 1997). The enormous growth of collaboration among nations and research institutions witnessed during the last decades is a function of science policy initiatives and the internal dynamics of science as well. Some authors (Nguyen, Ho-Le, and Le 2017) have even stated that international collaboration in research is now the norm rather than an exception.

The most common way to study collaboration is through co-authorship (Melin and Persson 1996). Numerous studies have pointed out that internationally collaborated articles have a higher scientific impact compared to domestic ones. For example, this has been seen in cases of Slovenia (Mali, et al. 2016), Spain (Bordons et al. 2014), and the United Kingdom (Adams 2013). Kremer, Werner (2009) and Hirv (2018) saw the same trend in Estonian, especially in natural sciences. Also, Estonian high rates of participation in European Union Framework Programmes (Ukrainski, Masso, and Kanep 2014) emphasize the importance of science internationalization, especially for a small country.

It is found (Lauk and Allik 2018) that Estonia managed to improve its scientific impact significantly during 2007-2018 despite low R&D funding. Estonian remarkable scientific progress also got mentioned by Schiermeier (2019) in Nature's article where it is underlined that ecology, molecular biology and genetics are doing especially well. Lauk and Allik (2018) call this phenomenon the puzzle of Estonian science because factors behind the increase are largely unknown.

The current article looks at Estonian international collaboration through a bibliometric perspective to understand how collaboration affects scientific output in terms of articles published, and citations received per article in Estonia during 2005–2015. For example, how much from a bibliometric perspective do large infrastructure-based collaborations (e.g. CERN and IDEFICS CONSORTIUM) influence Estonian scientific impact? Also, how does Estonia benefit from collaboration compared to other countries? Is it larger as MIRRIS Interim Report suggests which states that some countries like Estonia may have so-called *alibi position* in European Framework Programmes where project leaders believe that evaluators will have sympathy if they involve partners from small EU13 countries? These aspects are very important for policy implications because they could help to answer the puzzle of Estonian science. This article also gives an overview of which countries Estonian researchers' collaboration partners come from and which research fields have benefitted the most from international collaboration.

2. Theoretical background

The idea of bibliometric analysis has transformed from an intriguing possibility (de Solla Price 1962) into a regular tool for evaluation of the scientific quality of countries and institutions (Garfield 1979). The same also applies to Estonia. Eugene Garfield, the pioneer of scientometrics, states (Clarivate Analytics 2018, 2) "the total number of expressions [citations] is about the most objective measure there is of the materials importance to current research". Citations in scientometric analysis have three important roles (Layzell 1999): 1) they constitute a measurable objective for which resources are allocated; 2) offer a comparison between different projects based on previous results and costs; 3) enable reliable information to be independently audited. We use Griliches's (1979) knowledge production function (KPF) to describe the basic dynamics of scientific output and the importance of international collaboration. According to Griliches (1979), a scholar's scientific output Y is a function $F(X,K,u)$ where X stands for a current level of technological knowledge, *K* for capital, and *u* for unmeasured determinants of output. International collaboration plays here an important role because it gives an opportunity to combine these mentioned inputs (*X* and *K)* on an international level to increase the output of a researcher, as well as of the country.

When talking about output, scientific quality and impact from bibliometric perspective, usually means a number of citations an article receives. This understanding comes from the normative theory of citing behavior which is based on Robert Merton's sociological theory of science (Merton 1973). This theory states that scientists give credit to other scientists by citing publications they use and are influenced by. Unfortunately, in some cases, high citation counts do not always indicate quality. An article may be cited frequently because other authors are refuting its findings.

De Solla Price (1962) discovered that the growth of scientific papers is exponential. Because of this notion, we adapt Hayati and Didegah (2010) approach and use exponential growth function to determine growth rates of scientific articles. We compare growth rates of collaboration articles and domestic ones to see how the science systems in Estonia and other CEE countries are internationalizing in terms of co-authorship. The equation of this model is as follows in which *Y* is the number of co-authored papers, *b* is the coefficient of time called growth rate, and *t* is the independent variable (here time):

$$Y = a.e^{b.t} \tag{1}$$

It is stated that hyperauthorship in CERN strongly influences the co-authorship geography of small countries (Must 2014). There is no clear line of what is considered hyperauthorship. Some authors say that more than 50 authors in an article is an indication of hyperauthorship (Rousseau, Egghe and Guns 2018, 120, King 2013, 2) while others are more conservative and say this number is around 16 (Boffito et al. 2016, 1128, Gonzalez-Brambila, Veloso and Krackhardt 2013, Mohallem and da Fonseca 2015). Since the main purpose of this paper is the analysis of Estonian international collaboration and scientific impact, we cover both but focus more on conservative approach about hyperauthorship. The exclusion of highly collaborated papers restricts research to the papers that have a significant contribution from the (Mohallem and da Fonseca 2015) selected country's researchers. Gonzalez-Brambila, Veloso, and Krackhardt (2013) suggest that publications with a high number of co-authors can reflect another type of collaborative effort and not necessarily the 'actual' network embeddedness of researchers. Also, when the efforts are on a grander scale, it is doubtful that 100 or 50 researchers could have possibly written, edited, and approved the final work (Liesegang, Schachat, and Albert 2010).

3. Design of the study, data and methods

The metrics for the current study are retrieved using online InCites database; a research analytics solution provided by Clarivate Analytics (2018). InCites has the advantage of an extensive coverage of recognized, citation-based, and widely read scientific journals. The present study covers eleven years from 2005 to 2015. Although newer data is available, it is decided not to use the latest data because of the time dependence of citations (Wang 2013). For the basis of this analysis, the Essential Science Indicators (ESI) scheme from InCites database is selected to maintain the comparability of results to Lauk and Allik's (2018) study. Selected documents will only include research and review articles from Science Citation Index Expanded and Social Science Citation Index because only these two indexes are mapped to ESI. As a limitation of this schema, it excludes publications from Arts and Humanities, Conference Proceedings Citation Index, and Book Citation Index (Clarivate Analytics 2018). We use InCites advanced search options to determine international collaboration. Identification of collaboration is based on the article's address section. If there are two country addresses, then it is read as a product of international collaboration.

In the context of this article, we measure output in terms of scientific papers published and citations received by articles. It is important to make this distinction in measuring because single metrics cannot disentangle quality from quantity (Kaur, et al. 2015). For investigating scientific impact, we use Category Normalized Citation Impact (CNCI), an average percentile in the subject area (both provided by InCites), and also High Quality Science Index (HQSI). The CNCI of a paper is calculated by dividing the actual count of citing items by the expected citation rate for documents with the same document type, year of publication and subject area.

In order to compare countries' scientific excellence, we follow Allik's (2013) approach in calculating HQSI. The HQSI is computed for a country (or territory) as a sum of normalized scores of the mean impact (citations per paper) and the percentage of publications in the top 1% based on citations by category, year, and document type. Before summation, the number of citations per paper and the percentage of highly cited papers are transformed according to the following formula: $(X\text{-}M)/SD$ where M is the mean value of countries, X is a country's value, and SD is the standard deviation. Therefore, for computing this index, both components are first normalized (the mean of the transformed values equals zero, and standard deviation equals one) and then summed together. In order to avoid countries whose scientific influence is irrelevant, it is decided to include only those countries whose scientists were able to publish 4,000 or more papers over the observed eleven-year period.

For determining the most frequent research partners of Estonia, we use InCites analytic tools. In order to compare outcomes of collaboration, we use, in addition to previously mentioned metrics, scientific impact in terms of citations and average percentile in the subject area. Average percentile is calculated by dividing the actual count of citing items by the expected citation rate for documents with the same document type, year of publication and subject area (Clarivate Analytics 2018), and subtracting this result from 100. Subtraction is necessary to ensure that the results are as easy to interpret as possible because InCites by default uses 'inverted percentiles' (Bornmann 2013) where low percentile values mean high citation impact (and *vice versa*).

A simplified table with whom Estonia has collaborated does not give us the complete overview; therefore, in addition, the map of Estonian bibliometric network is visualized by means of VOSviewer. Also, for visualizing collaboration patterns, we use Excel radar graph with the InCites' Global Institutional Profiles Project (GIPP) scheme to determine collaboration patterns between countries and research areas. In order to figure out which research fields in Estonia have benefited the most from internationalization, we determine citation impact in percentiles for domestic and international collaboration articles in all ESI research fields.

4. Results

We see from Figure 1 that a number of Estonian collaboration articles has been steadily increasing and has multiplied compared to the beginning of the period. Range (1-16; 1-50; 1-5575) shows a number of collaboration articles based on only those articles where the numbers of authors stay between the mentioned range. All three lines run parallel until Estonia started getting more collaboration in consortia that created a divergence between the lines. These collaborations in consortia have also created a change in authorship dynamics – a small number of hyper-authored articles have a strong effect on co-authorship. Figure 2 shows how the average number of authors per article has changed over time. In order to deal with a large range of quantities, a logarithmic scale is used. Figure 2 shows that an average number of authors per paper has risen in all three cases, but for the interval 1-5575, the increase is enormous. Fluctuations in gray line (range 1-5575) may be due to different reasons: 1) Estonia started to participate in those consortia projects where fewer people were involved; 2) consortia, in general, have begun using fewer people; 3) not all people who participated are listed as authors.

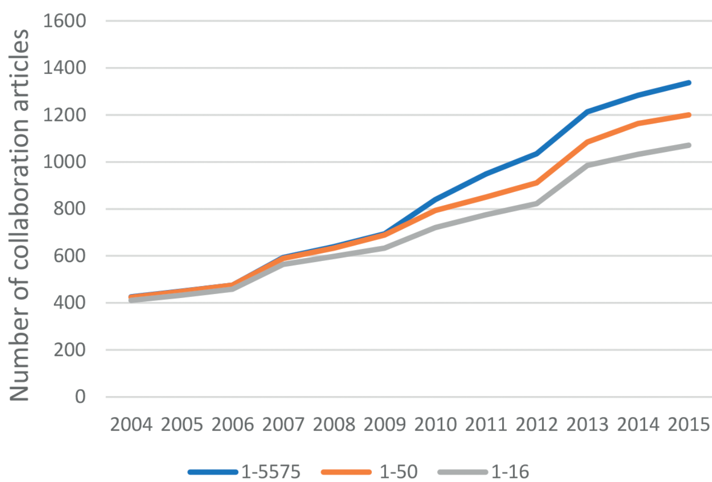

Figure 1. Number of collaboration articles.

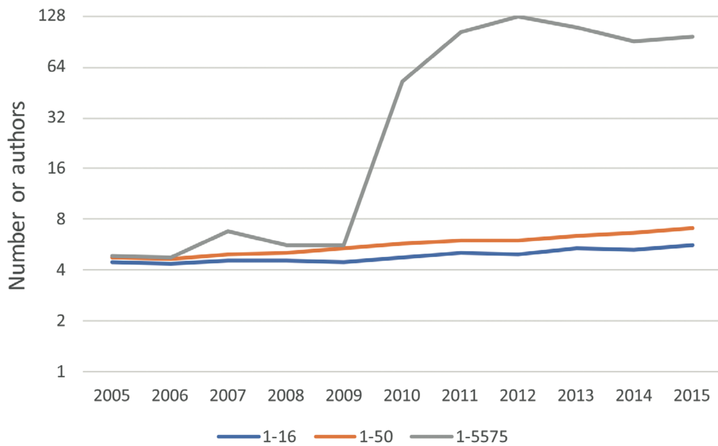

Figure 2. Average number of authors per article in logarithmic scale.

According to Web of Science (WoS) advanced search possibilities, most of the hyper-authored (more than 16 authors) articles are a result of collaborations in consortia, mostly in CERN (423) followed by IDEFICS CONSORTIUM (62) and EUROPEAN MALE AGEING STUDY GRP (48)) as shown in Table 1. WoS identified 865 group-authored articles and 347 names of different groups of authors, but a large number of those are name variations of the same groups (groups write their names slightly differently in different articles). Thirteen most frequent groups of authors are consolidated in Table 1. These numbers are robust estimates because InCites database does not fully support search filter for groups of authors and advanced search option in WoS is inaccurate because of naming misunderstandings. Also, a lot of groups do not mention their name at all in a paper. Depending whether a liberal or conservative approach about hyperauthorship is used, approximately 52-79% of hyper-authored articles mention consortium name in WoS.

Table 1. List of groups of authors

Rank	Groups of authors	Record count	% of 865
1	CERN	423	48.90%
2	IDEFICS CONSORTIUM	62	7.17%
3	EUROPEAN MALE AGEING STUDY GRP	48	5.55%
4	TOTEM COLLABORATION	14	1.62%
5-6	LIFELINES COHORT STUDY	13	1.50%
7	RIGHTTIMEPLACECARE CONSORTIUM	11	1.27%
8	EPICOM GRP	9	1.04%
9	EURO GBD SE CONSORTIUM	8	0.92%
10-13	CARDIOGRAM CONSORTIUM	7	0.81%
10-13	DIABIMMUNE STUDY GRP	7	0.81%
10-13	GIANT CONSORTIUM	7	0.81%
	Total	865	100%

It has been found earlier that Estonian scientific impact in terms of citations per paper has increased significantly (Allik 2015). We propose that Estonian high scientific impact depends heavily on hyper-authored articles. Must (2014) states that CERN collaborations strongly influence the co-authorship geography of small countries. She found that in small countries on average it affects the overall picture by about 44% in terms of co-authorship. We expect that in the analysis were hyper-authored articles are excluded Estonian scientific impact would drop significantly. Figure 3 shows that Estonian CNSI depends heavily on how much co-authorship is allowed in the analysis. In all cases, Estonian impact in terms of CNSI is around average (1) or higher. In later periods, allowing the maximum number of authors possible (1-5575) Estonian scientific impact doubles in terms of CNSI. Based on that we have to accept the suggestion that hyper-authored articles affect Estonian scientific impact and that should be taken into consideration when evaluating scientific results and research policy.

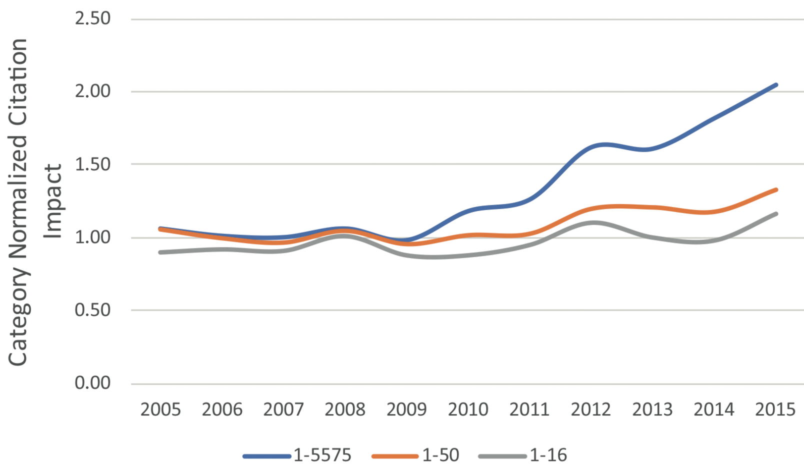

Figure 3. Estonian CNCI depending on co-authorship.

As described above, Estonian scientific impact is influenced by hyper-authored papers from 2010–2015. In order to determine in which research fields the positive effect of hyper-authored articles is the largest, the following figure (Figure 4) is constructed. It shows how much a range of authorship contributes to the total number of citations. At aggregated level articles with up to 16 authors account for 55% of total citations. This large difference comes from three research fields: *Molecular Biology & Genetics; Clinical Medicine; and Physics*. In the case *of Molecular Biology & Genetics*, articles with a number of authors up to 16 account for only 15,73% of total citations. It means that these highly collaborated articles where Estonian participation is small manage to generate around 84% of all citations in this research field. Articles with up to 16 authors account for 30% and 27% of total citations in *Clinical Medicine* and *Physics.* In the rest of the fields, this difference is smaller. When we are more liberal about what constitutes hyperauthorship and add articles where a number of authors stay between 17 and 50, we will see that the differences become smaller, but hyperauthorship (51-5575) still plays a major role.

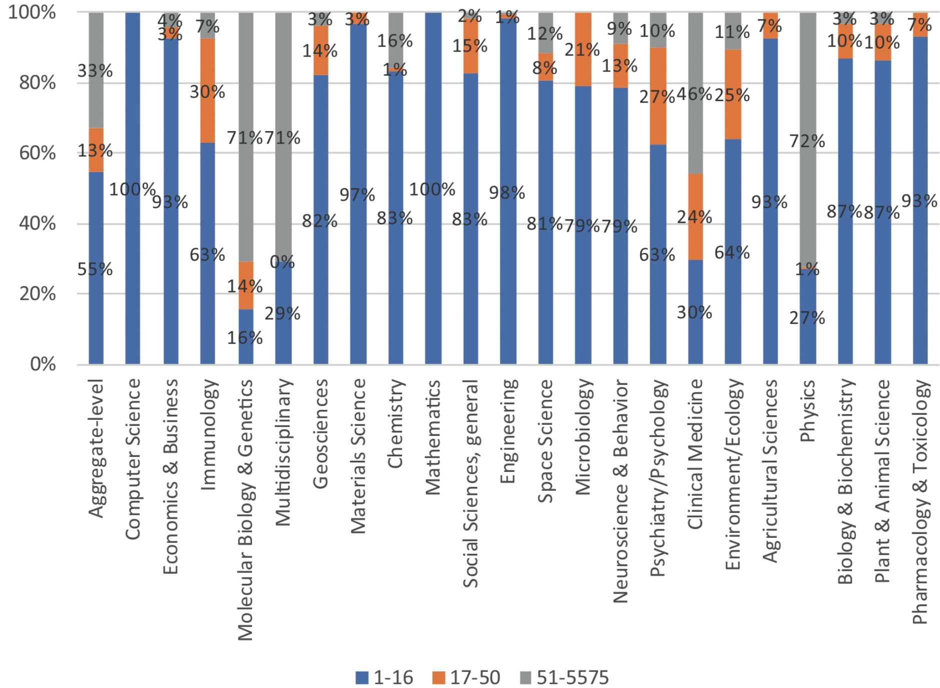

Figure 4. Influence of highly collaborated articles in different research fields (2010–2015).

Hyper-authored articles not only explain a large portion of the puzzle of Estonian science – how Estonian articles are cited 30% more frequently than average papers recorded by ESI (Lauk and Allik 2018) – but also Estonian high ranking on the High-Quality Science Index (HQSI) on Lauk and Allik study (2018). We see from Table 2 how the rank of HQSI depends on the range of co-authorship allowed in the analysis covering 2005–2015. In general, differences between different ranges measured by Spearman rank correlation are very low. For example, range 1-5575 has almost identical rank order ($r = 0.99$) compared to Lauk and Allik (2018) study. Range 1-5575 does not differ a lot from 1-50 ($r = 0.97$), but we see a larger difference ($r = 0.89$) when compared to 1-16 range. Although that correlation is still very high, some noticeable differences have emerged. For example, Iceland drops from the first position in scientific excellence to the twelfth position, Peru from sixth to thirtieth, and Estonia from twelfth to thirty-first.

Table 2. The countries HQSI rank depending on a range of co-authorship allowed in the analysis.

Name	1-16	1-50	1-5575	Lauk and Allik 2008-2018
SWITZERLAND	1	1	2	2
NETHERLANDS	2	3	4	4
DENMARK	3	2	3	5
SINGAPORE	4	7	8	6
USA	5	8	11	15
SCOTLAND	6	5	5	3
ENGLAND	7	6	7	9
UNITED KINGDOM	8	9	9	
HONG KONG	9	16	22	24
BELGIUM	10	10	10	10
SWEDEN	11	11	14	14
ICELAND	12	4	1	1
CANADA	13	13	20	18
WALES	14	12	13	8
AUSTRALIA	15	18	21	23
IRELAND	16	17	16	13
NORWAY	17	14	15	17
GERMANY (FED REP GER)	18	21	26	21
AUSTRIA	19	15	18	16
NORTHERN IRELAND	20	22	24	20
KENYA	21	19	19	22
FINLAND	22	23	23	19
FRANCE	23	24	31	26
ISRAEL	24	25	28	27
LUXEMBOURG	25	27	32	30
NEW ZEALAND	26	26	27	28
ITALY	27	28	35	32
SPAIN	28	33	38	35
SAUDI ARABIA	29	37	41	37

Name	1-16	1-50	1-5575	Lauk and Allik 2008-2018
PERU	30	20	6	7
ESTONIA	**31**	**29**	**12**	**12**
PORTUGAL	32	36	39	41
QATAR	33	34	30	42
TANZANIA	34	30	37	39
CYPRUS	35	38	25	31
GREECE	36	39	40	38
COSTA RICA	37	31	33	34
PHILIPPINES	38	32	29	25
UGANDA	39	35	36	33
CHINA MAINLAND	40	45	55	61
SOUTH AFRICA	41	40	44	43
URUGUAY	42	41	42	44
INDONESIA	43	42	46	52
JAPAN	44	48	56	55
UNITED ARAB EMIRATES	45	44	52	64
HUNGARY	46	43	43	40
THAILAND	47	47	51	59
SOUTH KOREA	48	55	62	63
SRI LANKA	49	50	34	29
CHILE	50	46	48	49
SLOVENIA	51	51	50	51
CZECH REPUBLIC	52	49	49	48
TAIWAN	53	59	65	70
MALAYSIA	54	61	70	71
ARGENTINA	55	56	58	57
GHANA	56	53	45	45
BANGLADESH	57	57	59	58
LEBANON	58	52	53	47
ETHIOPIA	59	62	64	66
VIETNAM	60	58	57	68
INDIA	61	67	80	84

Name	1-16	1-50	1-5575	Lauk and Allik 2008-2018
MEXICO	62	63	67	74
IRAN	63	75	84	88
COLOMBIA	64	60	47	54
KUWAIT	65	65	78	81
JORDAN	66	72	79	82
VENEZUELA	67	64	68	72
EGYPT	68	74	82	87
OMAN	69	66	69	62
CAMEROON	70	69	71	75
BRAZIL	71	71	81	85
LATVIA	72	54	54	56
ALGERIA	73	79	86	92
POLAND	74	68	75	78
SLOVAKIA	75	70	66	73
PAKISTAN	76	78	77	76
TURKEY	77	82	85	91
CROATIA	78	73	60	67
CUBA	79	77	73	79
MOROCCO	80	81	72	80
TUNISIA	81	85	89	96
BULGARIA	82	76	63	60
SERBIA	83	83	76	83
LITHUANIA	84	80	74	77
ROMANIA	85	84	83	86
IRAQ	86	86	87	94
RUSSIA	87	89	91	95
KAZAKHSTAN	88	88	92	97
NIGERIA	89	87	88	89
UKRAINE	90	90	90	93
BELARUS	91	92	61	69
ARMENIA	92	91	17	36

From this point, we continue only with a more conservative approach about hyperauthorship and limit co-authorship to sixteen because hyperauthorship strongly influences the scientific output of small countries. Because collaboration goes both ways, we propose that if we exclude large infrastructure-based collaborations, then there is a proportional relationship between national and collaboration citation impact. An increase in national citation impact leads to an equivalent increase in the collaboration citation impact. It is opposite to MIRRIS Interim Report (2014) which suggests Estonia is getting more from collaboration than other countries. It is well confirmed that collaboration increases general citations rates, but here we are interested in whether collaboration changes the proportion dynamics between countries. Correlation between gross national income *per capita* and scientific impact in terms of citation per paper is already confirmed (Allik 2013), and increasing the impact and effectiveness is one of the most common motives for collaboration (de Beaver and Rosen 1978).

Based on calculated HQSIs, international collaboration does not change the proportion dynamics between countries in scientific impact. Collaboration HQSI rank order replicates rank order from national articles by Spearman correlation of 0.83. It is understandable because collaboration goes both ways and in order to get something, participants have to give something equally valuable back. From Figure 5, we can see a proportional relationship between national and collaboration citation impact as proposed. *Y*-axis shows citations impact per paper for collaborated articles and *X*-axis for articles without collaboration. Estonia is slightly above the trend line, and it indicates that Estonia does not benefit more from international collaboration than other countries when hyper-authored articles are excluded. Countries who have a high scientific impact without collaboration will have an even higher impact with collaboration. Therefore, collaboration is not a mechanism for catching up to scientifically more advanced countries but is rather a mechanism for staying competitive because collaborated papers have on average higher scientific impact. It supports the argument that international collaboration in research is now the norm rather than an exception (Nguyen, Ho-Le and Le 2017). The only way countries can use collaboration as a catch-up mechanism is if they collaborate more on average than countries whom they are trying to catch up with.

If international collaboration is a way to compensate lack of resources does Estonia use it frequently enough to minimize the lag between it and more developed EU15 countries? As we can see from Table 3, the proportion of articles with international collaboration varies from country to county. Fifty percent of Estonian publications in 2005 were international collaboration articles. The comparison of the exponential growth rates of national and international publications shows that international collaboration is becoming more dominant in Estonia. The growth rate of domestic publications is 5.2% compared to 9.3 for collaboration articles, and 59% of publications in 2015 were collaboration articles. Growth rates show very strong path dependence (based on R^2), especially for international collaboration one, and it supports Arunachalam and Doss (2000) that in small countries international collaboration is necessary to overcome the lack of financial and human resources. In

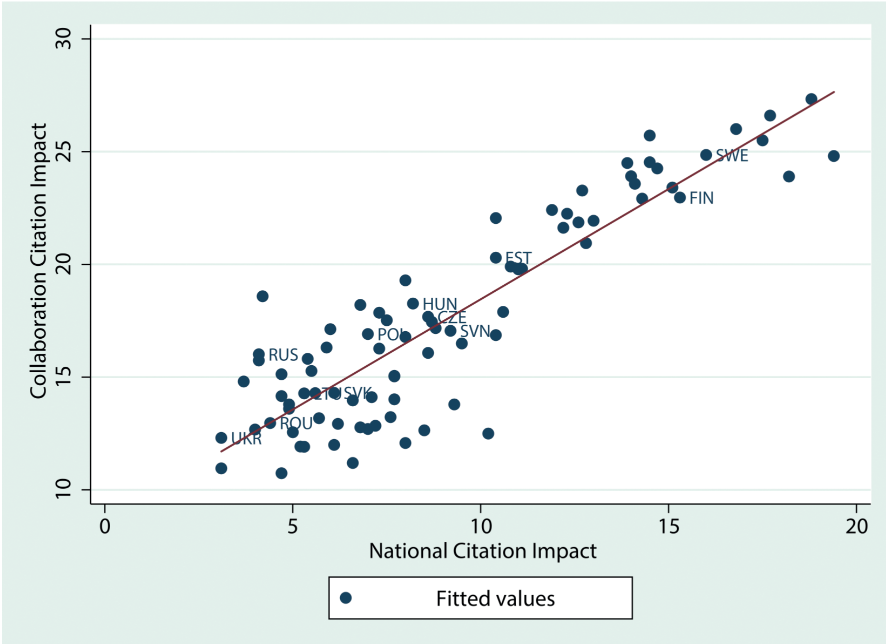

Figure 5. Proportional relationship between national and collaboration citation impact (2005–2015).

all observed countries, the growth of international collaboration articles follows more closely the exponential growth function than the growth of domestic articles. Lower R-squared in domestic output is possibly due to the volatility in local funding. The difference in exponential growth rates (collaboration-domestic) shows the current trends of internationalization. Although Estonia collaborates more internationally than the EU15 area and is pursuing collaboration as a strategy for catching up to advanced countries in terms of scientific impact, the gap in collaboration activity is slightly decreasing.

Table 4 on the next page shows which countries are the most frequent collaboration partners of Estonia. The table lists countries with whom Estonia has had at least 100 collaboration papers together. As shown in the table, the main partners of Estonian researchers were scientifically advanced countries with very high HQSI. All significant collaboration partners have HQSI above average and above Estonian 0.65. Although differences in citation impact in citations and average percentile are not large between different collaboration outcomes, there are significant differences in the proportion of articles in the top 1%. It seems participation of countries with very high scientific excellence (HQSI) is required to publish an article that gets into the 1%. In order to get a more sophisticated understanding of Estonian collaborations, the bibliometric co-authorship network is visualized.

Table 3. Exponential growth rates of international collaboration and domestic articles in 2005–2015

	Proportion of articles with collaboration in 2005 (%)	Domestic *Y*	R^2	Collaboration *Y*	R^2	Differenc e in growth rates (C-D)	Proportion of articles with collaboration in 2015 (%)
EU 15	32.0	$286831e^{0.0213x}$	0.80	$125963e^{0.0648x}$	0.99	4.3	42.3%
BULGARIA	53.8	$985.93e^{0.0332x}$	0.45	$1104.7e^{0.0107x}$	0.31	-2.25	49.2
CZECH REPUBLIC	42.6	$4341.4e^{0.0471x}$	0.84	$2834.2e^{0.0784x}$	0.98	3.13	50.7
ESTONIA	50	$481.31e^{0.0524x}$	0.71	$402.23e^{0.0933x}$	0.98	4.09	59.0
FINLAND	42.9	$5577.7e^{0.0062x}$	0.44	$3835.9e^{0.0678x}$	0.99	6.16	57.8
HUNGARY	47.6	$3194.9e^{0.0164x}$	0.53	$2694.3e^{0.0378x}$	0.97	2.14	51.2
LATIVA	58.1	$166.98e^{0.0984x}$	0.78	$184.48e^{0.0693x}$	0.88	-2.91	51.2
LITHUANIA	42.2	$885.32e^{0.0563x}$	0.36	$444.37e^{0.0664x}$	0.93	1.02	40.8
POLAND	35.3	$10593e^{0.0613x}$	0.95	$5288.9e^{0.0472x}$	0.92	-1.41	31.7
ROMANIA	46.1	$1984.3e^{0.1236x}$	0.69	$1322.2e^{0.0879x}$	0.95	-3.57	36.1
RUSSIA	34.2	$16949e^{0.033x}$	0.85	$8029.7e^{0.0232x}$	0.53	-0.98	32.2
SLOVAKIA	49.6	$1231.7e^{0.0375x}$	0.78	$1168.2e^{0.0488x}$	0.92	1.13	53.3
SLOVENIA	38.9	$1552.4e^{0.0404x}$	0.57	$812.56e^{0.0937x}$	0.96	5.33	50.4
SWEDEN	46.5	$10339e^{0.0152x}$	0.71	$8753e^{0.0681x}$	0.99	5.29	60.7

Table 4. Main bibliometric indicators of collaborations

Rank	Name	Collaboration partner`s HQSI	International collaborations	% of all collaboration articles	CNSI	Citation Impact	Average Percentile	% Documents in Top 1%
1	FINLAND	1.62	1485	21%	1.25	24.38	59.99	1.62%
2	SWEDEN	2.58	1297	18%	1.54	31.07	62.52	2.62%
3	USA	3.78	1114	16%	1.88	36.70	64.27	4.40%
4	GERMANY	2.13	1112	16%	1.55	28.55	62.49	1.80%
5	UK	3.44	1021	14%	1.68	33.16	66.39	3.23%
6	ENGLAND	3.59	837	12%	1.66	34.23	66.07	3.23%
7	ITALY	1.26	606	9%	1.98	34.94	67.72	3.80%
8	SPAIN	0.85	573	8%	2.08	35.84	68.80	3.32%
9	RUSSIA	-2.59	570	8%	0.77	16.00	47.63	0.70%
10	FRANCE	1.61	567	8%	1.77	34.57	67.13	2.82%
11	NETHERLANDS	4.07	393	6%	2.01	37.99	69.85	5.34%
12	POLAND	-1.72	387	5%	1.25	22.05	58.20	1.55%
13	DENMARK	3.87	385	5%	1.80	35.37	65.53	6.23%
14	NORWAY	2.14	355	5%	1.72	35.59	64.70	5.07%
15	BELGIUM	2.75	318	4%	1.85	28.01	68.48	3.46%
16	LATVIA	-1.67	315	4%	0.94	15.98	56.08	0.63%
17	LITHUANIA	-2.14	300	4%	1.01	18.91	56.69	1.33%
18	SWITZERLAND	4.72	264	4%	2.40	44.04	68.02	6.82%
19	CZECH REPUBLIC	-0.80	249	3%	1.43	22.62	61.31	2.81%
20	CANADA	2.48	246	3%	2.19	39.15	65.72	6.50%
21	SCOTLAND	3.60	228	3%	1.78	30.37	68.80	2.63%
22	AUSTRALIA	2.44	225	3%	2.30	50.79	69.18	7.11%
23	HUNGARY	-0.36	213	3%	1.54	24.10	65.84	1.41%
24	CHINA MAINLAND	0.07	189	3%	1.95	26.69	65.48	5.29%
25	JAPAN	-0.29	174	2%	1.55	26.10	59.48	2.87%

Bibliometric co-authorship network (Figure 6) shows that collaboration is not only between two countries and Estonia is embedded into a larger network. Colors used in the figure indicate how productive given collaborations have been in terms of normalized citations. The thickness of the lines between countries demonstrates the strength of a connection in terms of fractionalized co-publications. The social network analysis confirms the assumption that collaborations with authors from countries who have very high HQSI are necessary to write top cited articles. Collaborations with Russia, Ukraine, Lithuania, and Latvia have not increased Estonian citations impact as much as collaborations with scientifically advanced countries. We also see that collaborations which involve a wider specter of countries produce articles with higher impact (it is evident in the number of lines that connect countries). Estonian largest collaboration partners in this social network are brokers by providing linkages to other countries.

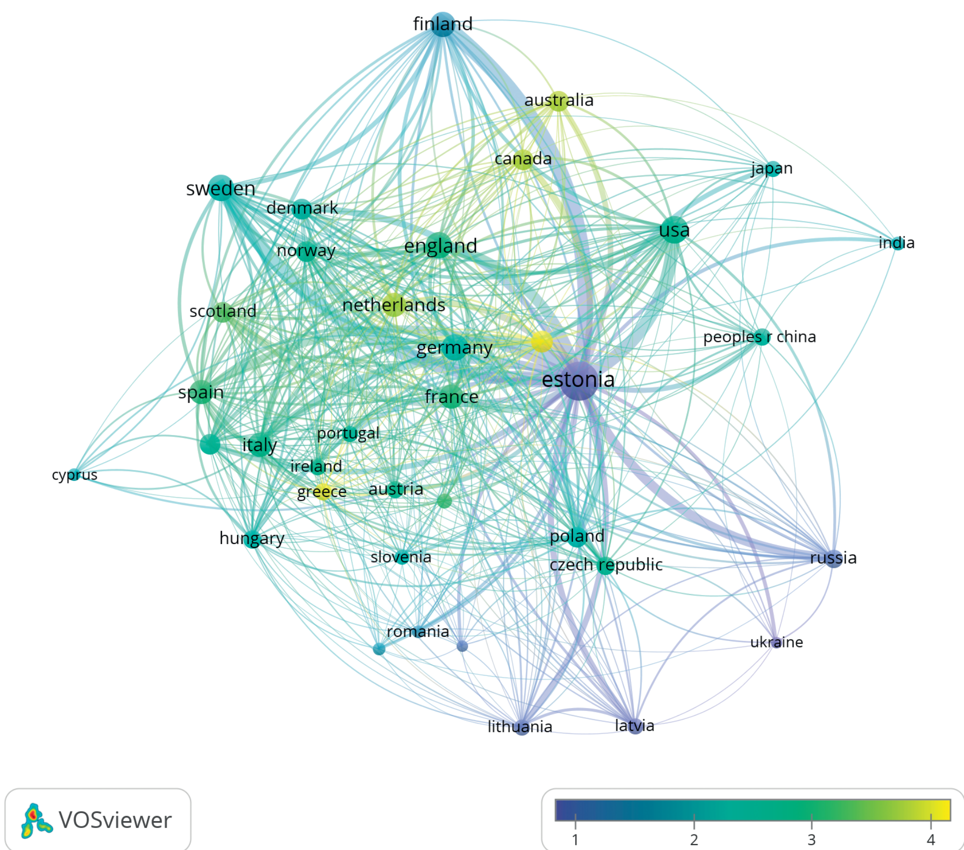

Figure 6. Visualization of Estonian bibliometric network.

Results also show (Figure 7) that the frequency of a country varies by disciplines. Estonian largest collaboration partner Finland is represented in all research domains equally just like the United States. Collaborations with Germany are less often in *Social Sciences* compared to other research areas. Collaborations with Sweden are the most frequent in *Life Sciences* and especially in *Clinical, Pre-Clinical & Health.*

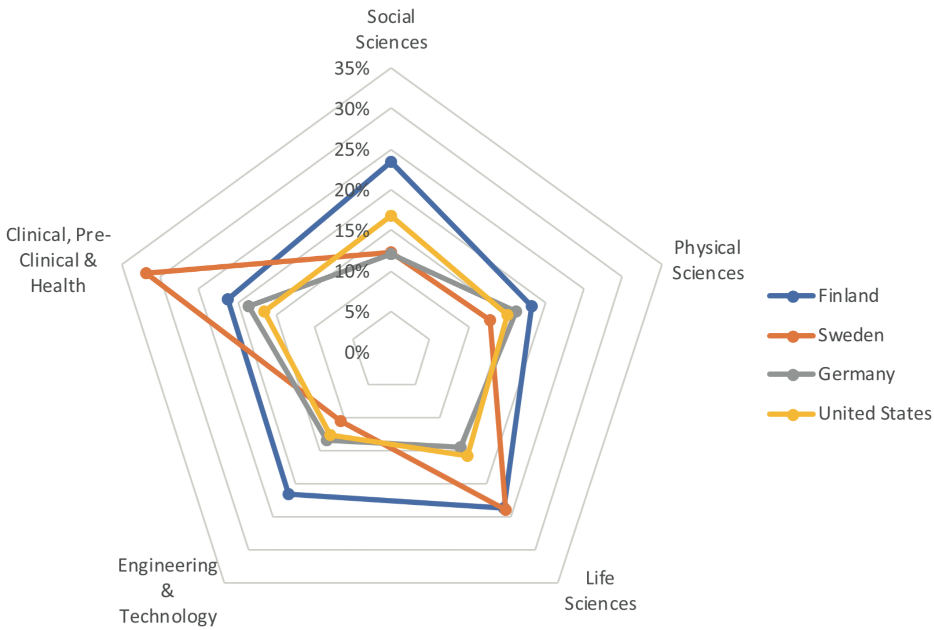

Figure 7. Frequency of collaborations in various disciplines

In Figure 8 we can see which research fields have been most active in international collaboration and which ones have benefitted the most. 72 percent of publications in *Immunology* and 69 percent of publications in *Molecular Biology & Genetics* have international collaboration. *Mathematics* (32%) and *Economics & Business* (28%) have the lowest proportion of international co-publications. The biggest beneficiary from internationalization is *Space Science* – the difference between international and national articles is 29 percentiles. *Space Science* is followed by *Immunology* (21*), Psychiatry/Psychology (*19), *Pharmacology & Toxicology* (18*)*, and *Microbiology* (12). The difference is the smallest in *Materials Science (3), Mathematics (5),* and *Geosciences* (5). These findings indicate that benefit from collaboration is volatile, and all research fields have benefitted.

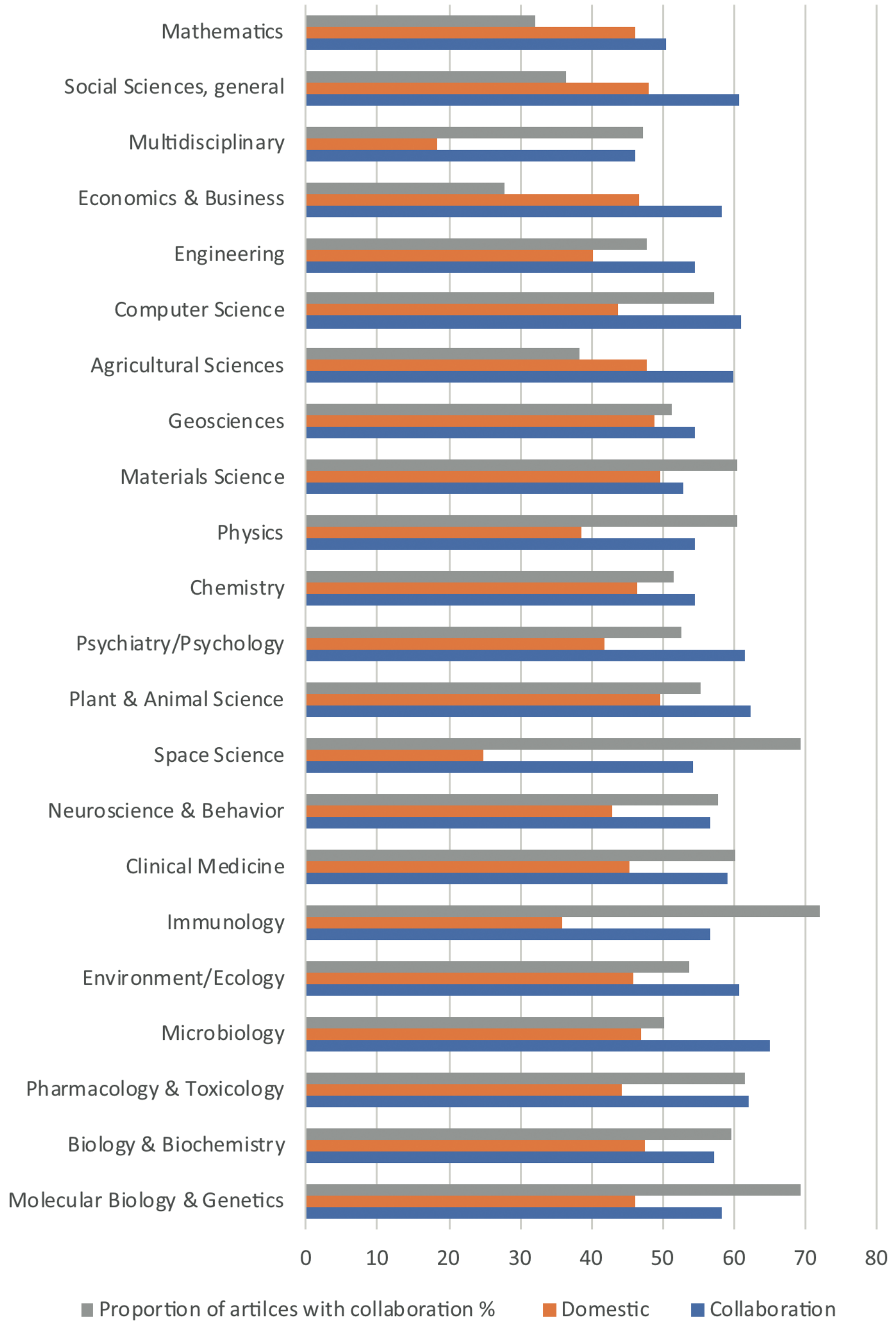

Figure 8. The distribution of publications with domestic authorship and international collaboration in different research fields.

5. Conclusions and implications

The current study used a bibliometric analysis to examine the relationship between collaboration and the quantity and quality of Estonian research publications. Results show that international collaboration has an important role in determining Estonian scientific impact. Estonian impact in terms of citations has risen especially due to participations in international research consortia. Collaborations in consortia (especially CERN) enable Estonian scientists to participate in high-end science projects and allow Estonian scientists to get valuable experiences. On the other hand, these collaborations have created a challenge from an evaluation viewpoint. These hyper-authored collaborations have doubled Estonian scientific impact during 2010–2015. This is a large difference, and it can cause misleading results and conclusions in science evaluation. For example, participations in hyper-authored articles explain a large proportion of the puzzle of Estonian science – how it managed to improve scientific impact significantly despite low R&D funding as stated by Lauk and Allik (2018).

One possible way to deal with hyperauthorship is to limit the number of authors allowed in the analysis. When efforts are on a grander scale, it is doubtful that all researchers could have possibly have written, edited, and approved the final work. A number of co-authors allowed in an analysis can vary but, in this article, we followed both liberal (up to 50 authors) and conservative (up to 16 authors) approach but focused more on conservative one. Articles limited to conservative approach at the aggregated level account only for 55% of total citations. This large difference comes from three research fields: *Molecular Biology & Genetics; Clinical Medicine;* and *Physics*. As a policy recommendation, the mentioned research fields should be paid close attention because scientific impact in terms of citations depends strongly on large infrastructure-based collaborations. In the case of *Molecular Biology & Genetics*, conservative approach about co-authorship accounts for 16% of total citations. It means that 84% of citations in this research field come from international co-publications where Estonian participation is rather small. Under the liberal approach about what constitutes hyperauthorship, the difference is slightly smaller, but hyperauthorship still plays a major role.

The scientific quality (HQSI) rank order of countries based on internationally collaborated articles replicates a rank order from national articles by Spearman correlation of 0.83. A benefit from international collaboration in terms of citations follows a proportional relationship between national and collaboration citation impact. Based on this finding it is necessary for Estonia to contribute proportionally more in terms of R&D in order to catch up with more advanced countries. Investment has a multiplicative effect because it enables getting better collaboration opportunities. Collaboration goes both ways, and in order to get something, participants have to give something equally valuable back. It indicates that international collaboration in most cases is not a mechanism for catching more advanced countries *per se* but is rather a mechanism of staying competitive because the one who is not collaborating, is the one who is left behind.

Because collaboration articles are more cited, countries can use collaboration as a mechanism to catch up with scientifically more advanced countries only if they collaborate more than they do. Currently, Estonia is pursuing this strategy but struggles to keep up with growing financing needs. When fifty percent of Estonian publications in 2005 were collaborated internationally, then in 2015 it was already 59%. The growth rate of a number of domestic publications was 5.2% compared to 9.3 for collaboration articles. An increasing internationalization trend supports the argument that in small countries international collaboration is necessary to overcome the lack of financial and human resources.

The frequency of Estonian collaboration partners varies by discipline, but the main partners were scientifically advanced countries with high scientific capabilities. Although differences in citation impact are not large between different collaboration outcomes, there are significant differences in the proportion of articles in the top 1%. Results show that participation of countries with very high scientific excellence is required to publish an article that gets into the 1%. Also, these largest collaboration partners in Estonian science network are brokers by providing linkages to other countries. Collaboration with researchers from countries such as Russia, Ukraine, Lithuania, and Latvia has not increased Estonian citation impact as much as collaborations with scientifically advanced countries. If hyperauthorship is excluded, the biggest gain in scientific impact from internationalization over the observed period was seen in *Space Science* – the difference between international and domestic articles was 29 percentiles. Space Science was followed by *Immunology, Psychiatry/ Psychology, Pharmacology & Toxicology,* and *Microbiology*.

Here is an important challenge for policymakers – the impact of collaboration on research is positive and significant, its outcome in terms of visibility is greater than transaction costs (coordination costs of collaborative research projects of collaboration). Strategic questions are whether there are such consortia in every research area, and whether to invest in it or competition-based collaborations. Besides consortia, competition-based collaborations are an option worth considering because previous results have shown that when hyperauthorship is left out, a combination of EU funding and international collaboration produces the most cited scientific articles (Hirv 2018).

Although Estonia collaborates more internationally than EU15 area and therefore uses collaboration as a mechanism of staying competitive or even for catching up with more advanced countries, the gap in collaboration activity (as measured by co-publications) is slightly decreasing. This is not a problem because 'impact' in scientometrics seems to undergo a taxonomic change, where the impact is no longer defined as the impact on science alone (measured by citations), but on all sectors of society (e.g. culture, economics, or politics) (Bornmann and Haunschild 2016). Therefore, Estonia needs to focus more on other aspects of science in addition to the impact measured by citations.

Address:
Tanel Hirv
Faculty of Economics and Business Administration
University of Tartu
J. Liivi tn 4
51009 Tartu, Estonia
E-mail: tanel.hirv@ut.ee and tanelhirv@gmail.com
Tel.: +372 56833028

References

Adams, Jonathan (2013) "Collaborations: the fourth age of research". *Nature* 497, 7451, 557–560.

Allik, Jüri (2013) "Factors affecting bibliometric indicators of scientific quality". *Trames* 17, 3, 199–214.

Allik, Jüri (2015) "Progress in Estonian science viewed through bibliometric indicators (2004–2014)". *Proceedings of the Estonian Academy of Sciences* 64, 2, 125–126.

Arunachalam, S. and M. J. Doss (2000) "Science in a small country at a time of globalisation: domestic and international collaboration in new biology research in Israel". *Journal of Information Science* 26, 1, 39–49. doi:10.1177/016555150002600104.

Boffito, Daria C., Christian A. Patience, Paul A. Patience, François Bertrand, and Gregory S. Patience (2016) "How do you write and present research well? 8 – Assign authorship according to intellectual involvement". *Canadian Journal of Chemical Engineering* 94, 6, 127–1134. doi:10.1002/cjce.22479.

Bordons, Marıa, Borja Gonzalez-Albo, Javier Aparicio, and Luz Moreno (2014) "The influence of R&D intensity of countries on the impact of international collaborative research: evidence from Spain". *Scientometrics* 102, 2, 1385–1400. doi:10.1007/s11192-014-1491-4.

Bornmann, Lutz (2013) "How to analyze percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes, and top-cited papers". *Journal of the Association for Information Science and Technology* 64 ,3. 587–595. doi:10.1002/asi.22792.

Bornmann, Lutz and Robin Haunschild (2016) "Does evaluative scientometrics lose its main focus on scientific quality by the new orientation towards societal impact?" *Scientometrics* 110, 2, 937–943. doi:10.1007/s11192-016-2200-2.

Clarivate Analytics (2018) Essential Science Indicators Research Areas. Available online at <http://ipscience-help.thomsonreuters.com/inCites2Live/filterValuesGroup/researchAreaSchema/esiDetail.html>.

Clarivate Analytics (2018) *InCites indicators handbook*. Clarivate Analytics. Available online at <http://ipscience-help.thomsonreuters.com/inCites2Live/8980-TRS/version/default/part/AttachmentData/data/InCites-Indicators-Handbook-6%2019.pdf>. Accessed on 07 May 2019.

de Beaver, D. B. and R. Rosen (1978) "Studies in scientific collaboration. Part 2: Scientific co-authorship, research productivity and visibility in the French scientific elite, 1799–1830". *Scientometrics* 1, 2, 133–149.

de Solla Price, Derek J. (1962) *Little science, big science – and beyond.* New York: Columbia University Press.

Garfield, Eugene (1979) *Citation indexing: its theory and application in science, technology, and humanities*. New York: John Wiley & Sons.

Gonzalez-Brambila, N. Claudia, Francisco M. Veloso, and David Krackhardt (2013) "The impact of network embeddedness on research output". *Research Policy* 42, 9, 1555–1567. doi:10.1016/j.respol.2013.07.008.

Griliches, Zvi (1979) "Issues in assessing the contribution of research and development to productivity growth". 10,1, 92–116.

Hayati, Zouhayr and Fereshteh Didegah (2010) "International scientific collaboration among Iranian researchers during 1998-2007". *Library Hi Tech* 28, 3, 433–446. doi:10.1108/07378831011076675.

Hirv, Tanel (2018) "Effects of European Union funding and international collaboration on Estonian scientific impact". *Journal of Scientometric Research* 7, 3, 181–188.

Katz, J. Sylvan and Ben R. Martin (1997) "What is research collaboration?". *Research Policy* 26, 1, 1–18. doi:10.1016/S0048-7333(96)00917-1.

Kaur, Jasleen, Emilio Ferrara, Filippo Menczer, Alessandro Flammini, and Filippo Radicchi (2015) "Quality versus quantity in scientific impact". *Journal of Informetrics* 9, 4, 800–808. doi:10.1016/j.joi.2015.07.008.

King, Christopher (2013) "Multiauthorpapers: onward and upward". *Annual Report, Thomson Reuters* 4. Available online at <http://archive.annual-report.thomsonreuters.com/2012/_files/pdf/MultiauthorPapers_ChrisKing.pdf>. Accessed on 07 May 2019.

Kremer, Reinhard K. and Marx Werner (2009) "Aspects of the scientific cooperation of Estonia and Germany in view of bibliometry". *Proceedings of the Estonian Academy of Sciences* 58, 4, 255–262. doi:10.3176/proc.2009.4.07.

Lauk, Kalmer and Jüri Allik (2018) "A puzzle of Estonian science: how to explain unexpected rise of the scientific impact". *Trames*22, 4, 329–344. doi:10.3176/tr.2018.4.01.

Layzell, D. T. (1999) "Linking performance to funding outcomes at the state level for public institutions of higher education: past, present, and future". *Research in Higher Education* 40, 2, 233–246. doi:10.1023/A:1018790815103.

Liesegang, Thomas J., Andrew P. Schachat, and Daniel M. Albert (2010) "Defining authorship for group studies". *American Journal of Ophthalmology* 150, 2, 135–137. doi:10.1016/j.ajo.2010.06.001.

Mali, Franc, Toni Pustovrh, Rok Platinovšek, Luka Kronegger, and Anuška Ferligoj (2016) "The effects of funding and co-authorship on research performance in a small scientific community". *Science and Public Policy* 44, 1, 486–496. doi:10.1093/scipol/scw076.

Melin, G. and O. Persson (1996) "Studying research collaboration using co-authorships". *Scientometrics* 36, 3, 363–377.

Merton, Robert King (1973) *The sociology of science: theoretical and empirical investigations*. Chicago: University of Chicago Press.

Mohallem, José R. and Norma E. da Fonseca (2015) "Brazilian impact factor of physics journals – the third side of the coin". *Anais da Academia Brasileira de Ciências* 82, 2, 1233–1238. doi:10.1590/0001-3765201520140294.

Must, Ülle (2014) "The impact of multi-authored papers – the case of a small country". *Journal of Scientometrics and Information Management* 8, 1, 41–47. doi:10.1080/09737766.2014.916874.

Nguyen, Tuan V., Thao P. Ho-Le, and Ut V. Le (2017) "International collaboration in scientific research in Vietnam: an analysis of patterns and impact". *Scientometrics* 110, 2, 1035–1051. doi:10.1007/s11192-016-2201-1.

Rousseau, Ronald, Leo Egghe, and Raf Guns (2018). *Becoming metric-wise: a bibliometric guide for researchers*. Oxford: Chandos Publishing.

Saublens, Christian (2014) *Participation of EU13 countries in FP7: Mobilising Institutional Reforms for Better RDI Systems/ Institutions in Europe*. Part 1: *MIRRIS*, 200.

Schiermeier, Quirin (2019) "How Estonia blazed a trail in science – a small nation found strength in research after joining the European Union". *Nature* 565, 416–418.

Ukrainski, Kadri, Jaan Masso, and Hanna Kanep (2014) "Cooperation patterns in science within Europe: the standpoint of small countries". *Scientometrics* 99, 3, 845–863.

Wang, Jian (2013) "Citation time window choice for research impact evaluation". *Scientometrics* 94, 3, 851–872. doi:10.1007/s11192-012-0775-9.