CWTS Meaningful metrics

BIBLIOMETRIC REPORT

Bibliometric impact analysis of the Academy of Finland's Centre of Excellence Programmes

December 15th, 2015

Universiteit Leiden

December 15th, 2015
CWTS B.V.
Centre for Science and Technology Studies,
Leiden University

Bibliometric impact analysis of the Academy of Finland's Centre of Excellence Programmes

For: Academy of Finland (AKA)

Anu Nuutinen, *Senior Science Adviser*

Maiju Gyran, *Science Adviser*

Tel. +35 829 5335 015

E-mail anu.nuutinen@aka.fi; maiju.gyran@aka.fi

Project team

Joost Kosten MSc, *Project leader*

Ludo Waltman PhD, *Senior researcher/project consultant*

CWTS B.V.

P.O. Box 905

2300 AX Leiden, The Netherlands

Tel. +31 71 527 6104

Fax +31 71 527 3911

E-mail kostenmjf@cwts.leidenuniv.nl

Data analysis: CWTS, Leiden University; data source: Thomson Scientific/ISI

PDF: The report is available at www.aka.fi.

December 15th, 2015
CWTS B.V.
Centre for Science and
Technology Studies,
Leiden University

Table of Contents

Executive summary

In this report the results of a bibliometric analysis of the publications by the Finnish Centres of Excellence published during the period 2006-2013 are presented. The data collection has been performed by the Academy of Finland on the basis of the reporting system used by the Centres of Excellence. CWTS has matched these publications with the database. Approximately 7,000 publications indexed in the Web of Science (WoS) database have been included in the bibliometric analysis.

The publication output has increased slightly. This can be attributed to the fields Biomedical and health sciences and Life and earth sciences. The citation impact shows a little decrease. The majority of CoE publications have been published in the fields Biomedical and health sciences and Physical sciences and engineering. Relatively the field Life and earth sciences becomes more important while the field Physical sciences and engineering becomes less important. Publication output is low in the fields Mathematics and computer science and Social sciences and humanities. These are also the fields with a moderate internal coverage, indicating that scientific communication outside the WoS realm may be important in those fields. The citation impact decreases substantially only in the field Physical sciences and engineering. Publications by the CoEs have a higher citation impact than other publications with (a) Finnish affiliation(s), but the citation impact of all Finnish publications is hardly influenced by the citation impact of the relatively small set of publications by CoEs.

International collaboration dominates the CoEs collaboration profile, both in terms of (relative) output and citation impact. The citation impact of publications not resulting from international collaboration is considerably lower and has clearly decreased. At the level of main fields, the collaboration profile of the CoEs shows much diversity. International collaboration is important in terms of output and citation impact in any field, but the role of national collaboration and non-collaboration differs per field. Industrial collaboration plays a marginal role in publication output by CoEs but has increased in all five fields. The citation impact seems to be higher for publications resulting from industrial collaborative publications as compared with publications not resulting from industrial collaboration, but this differs per field. The share of industrial collaborative publications is comparatively high in Mathematics and computer science and Physical sciences and engineering.

1. Introduction

The Academy of Finland's Centres of Excellence (CoEs) play a prominent role in the Finnish research environment. The Centre of Excellence programme is designed to strengthen the international competitiveness of Finnish research and to increase its visibility and appreciation. In addition, the programme aims at developing creative and productive research and training environments that generate excellence and contribute to the overall progress of science and society. The CoE programme facilitates the establishing of research consortia between different fields of science and research. It promotes the efficient use of research infrastructures and opens up new opportunities for increased national and international cooperation. Furthermore, the CoE programme creates favourable conditions for scientific breakthroughs and in this way stimulates the regeneration of science and research. As a whole, the CoE programme supports the development of the Finnish research system.

Funding is provided for a six-year term. With the long-term funding CoEs can make long term plans and take risks in their research lines. CoEs are co-funded from the Academy of Finland, universities, research institutes, business sector and many other national and international funding sources. The key criteria for selecting CoEs are the scientific quality and innovativeness of their research. Other criteria include the feasibility of the research plan, the qualifications of team members, networking contacts and the contribution to the promotion of professional research careers and researcher training.

The bibliometric analysis carried out by CWTS Leiden is part of the final evaluation of the CoE Programmes over the years 2006-2011 and 2008-2013. The CoE Programme for 2006–2011 included 23 units with the total funding of 63.1 million euros from the Academy of Finland. The CoE Programme for 2008–2013 included 18 units and the Academy's funding was 56.3 million euros. This report aims to help answering two questions posed by the Academy of Finland about the publication activity and citation impact of CoEs in 2006-2013:

1. What is the publication output, citation impact and collaborative profile of CoEs?
2. What is the impact of CoE publications on the level of Finnish science?

In order to answer these questions, a diversity of bibliometric aspects related with the publications by the CoEs and by Finland as a whole are presented. Two periods of analysis are relevant: publications published during 2006-2009 (with citations included until 2010) and during 2010-2013 (with citations included until 2014). In this report, the methodological approach is very similar to the CWTS Leiden Ranking methodology. This enables the comparison with the CWTS Leiden Ranking results. In this way, there are 750 benchmark institutions available at www.leidenranking.com.

Results are presented both at the aggregate level and at the level of the five main fields as defined by the CWTS Leiden Ranking 2015. The five main fields are: Biomedical and health sciences, Life and earth sciences, Mathematics and computer science, Physical sciences and engineering, and Social sciences and humanities. Chapter 2 presents the Web of Science database, the data collection methods and a brief description of the bibliometric indicators. In Chapter 3, the results of a general bibliometric analysis of the CoEs are presented. Moreover, since not all publications could be included in the analysis based on fractional counting, a brief analysis of the excluded publications has been included. Chapter 4 focuses on differences between collaboration types in terms of output and citation indicators. This chapter also includes analysis of an industrial collaboration. In Chapter 5, a number of conclusions are drawn based on the analysis in Chapters 3 and 4.

2. Data collection and indicators

Data collection is a crucial step in any bibliometric analysis. In this chapter, the database, the method of data collection, the main characteristics of the dataset, and the bibliometric indicators are briefly introduced.

2.1. Database

Our CWTS Citation Index (CI) system has been used for the data collection and bibliometric analysis. The core of this system comprises of an enhanced version of the Thomson Reuters Web of Science (WoS) database. Within this database, the following citation indices are used: Science Citation Index Expanded (SCIE); Social Sciences Citation Index (SSCI) and Arts & Humanities Citation Index (A&HCI). We note that our in-house version of the WoS database includes a number of improvements over the original WoS database. Most importantly, our database uses a more advanced citation matching algorithm and an extensive system for address unification. We therefore calculate our indicators based on our in-house version of the WoS database.

WoS is a bibliographic database that covers the publications of about 12,000 journals in the sciences, the social sciences, and the arts and humanities. Each publication in WoS has a document type. The most frequently occurring document types are ‘article’, ‘book review’, ‘correction’, ‘editorial material’, ‘letter’, ‘meeting abstract’, ‘news item’, and ‘review’. Article, review, and letter are the document types that are most often included in a bibliometric analysis. For this analysis, letters have been excluded due to their limited scientific importance.

2.2. Counting method

When analyzing the bibliometric research output of an entity, publications can be counted fully or fractionally. In case of full counting, publications are fully assigned to an institute or country, irrespective of the number co-authors from different institutes or countries. If publications are fractionally counted, publications are only partly assigned in case of co-authorship. In this way, the effect of scientific collaboration is leveled out. Fractional counting is possible at different levels, e.g. countries or institutes. In this report, the publication output has been counted fractionally at the level of institutes in a way which best corresponds to the fractional counting method used in the CWTS Leiden Ranking. For this analysis, the share of all Finnish addresses on a publication has been assigned to the Centre of Excellence to which the publication belongs. We have applied this method since we cannot distinguish the addresses belonging to the Centres of Excellence from addresses of other organizations in Finland. This means that publications are counted fractionally regarding the international collaborative aspect, but are still fully counted as far as national collaboration is involved (see paragraph 4.1 for a definition of national and international collaboration). Since some publications by Centres of Excellence do not have Finnish addresses, these publications cannot be included in an analysis based on fractional counting.

2.3. Data collection

The data has been collected in two ways. For the analysis of Centres of Excellence, a list of publications has been provided by the Academy of Finland. This list has been compiled by the Academy of Finland in the following way. The CoEs report their staff and outputs (e.g. publications, completed degrees) annually. Over the eight-year period the Academy of Finland's reporting system has changed. All publication data that the CoEs had submitted to either of the reporting systems has been collected and combined. The original publication dataset consists of the whole publication data reported by the CoEs to the Academy of Finland. Any additional publication data has not been collected.

The bibliographic information in the whole dataset has been reorganized into a machine-readable format. This part of the data collection was very laborious because it had to be done mainly manually, publication by publication. Publications such as scientific books and book chapters which are not covered by the Web of Science were not reorganized into a suitable format for the matching process.

After an initial matching attempt, CWTS produced a data set of non-matched publications which, however, had a good likelihood to be matched with updated or more complete bibliographic information. These publications included, for example, journal articles which had originally been reported as 'submitted' or 'accepted' for publication. The Academy of Finland updated the bibliographic data with the help of the online Web of Science database.

CWTS has matched this final publication set with the CWTS version of the Web of Science database. Matching has taken place on the basis of, amongst others, author, title, publication year, and journal title.

The second approach concerning the data collection is based on the use of address information on publications in order to collect data for the country level analysis. All publications which include authors with an institutional affiliation in Finland have been included for the country level analysis.
For the analysis of Finland excluding the publications by Centres of Excellence, both data collection methods have been combined. First, all publications which contain author affiliations with a Finland based address have been collected and then all publications from the Centres of Excellence set have been excluded from the set.

The Centres of Excellence data and the publications data for Finland have been collected for the period 2006 until 2013. The analysis has been conducted for the publication windows 2006-2009 and 2010-2013 for the following units of analysis: Centres of Excellence, Finland including CoEs, and Finland excluding CoEs.
The citation windows include one additional year after the publication window ends. For instance, in the case of publications from 2006-2009, citations have been counted until the end of 2010.

Only part of the publications which could be matched with the Web of Science database have been included in this analysis. The following criteria should be met by a publication to be included in the analysis. The publication

1. is a WoS core publication; and
2. has been assigned to a sub-field (required for normalization); and
3. has at least one Finnish address (only relevant for fractional counting).

CWTS has defined core publications on the basis of characteristics of publications. A core publication is a non-retracted article or review written in English of which the authors are non-anonymous and which has been published in a core journal. A journal is classified as a core journal if it meets two criteria. First, it should have an international scope, which is determined on the basis of the countries in which the authors publishing in the journal or citing to the journal reside. Second, the journal should have a sufficiently large number of references to other core journals. A list of core and non-core journals is available on the CWTS Leiden Ranking website.

Table 2.1 presents an overview of the sets of publications which have been analysed. The column documents indicates the number of underlying publications. In case of full counting, the number of documents equals the total number of publications. The number of documents is larger for the CoEs in case of full counting, because some CoE publications had to be excluded from the analysis based on fractional counting since there these publications did not have a Finnish address among the listed affiliations.

Table 2.1 Final dataset for bibliometric analyses of the CoEs.

Unit of analysis	Counting method	Articles	Reviews	Total Publications	Documents
CoEs (2006-2013)	Full	6,787	312	7,099	7,099
CoEs (2006-2013)	Fractional	4,485.5	204.1	4,689.6	6,836
Finland incl. CoEs (2006-2013)	Fractional	50,554.3	2,271.7	52,825.9	77,847
Finland excl. CoEs (2006-2013)	Fractional	46,068.7	2,067.5	48,136.3	71,011

2.4. Bibliometric indicators overview

The indicators below are grouped by performance dimensions. Self-citations are excluded before impact indicators are calculated. More detailed information is provided in Appendix I and Appendix II.

In this report, the following indicators will be provided for each unit of analysis: P, MCS, TCS, MNCS, MNJS, PP(top 1%), PP(top 10%), P(top 10%), PP(top 50%), PP(uncited), PP(self-citations), PP(collab), PP(int collab) and Internal coverage. The other indicators can be obtained by the combination of some of the given indicators (e.g. TNCS = MNCS * P; TNJS = MNJS * P, P(uncited) = PP(uncited) * P, etc.).

Table 2.2. Overview of CWTS bibliometric indicators.

Indicator	Dimension	Definition
P	Output	Total number of publications.
TCS	Impact	Total number of citations
MCS	Impact	Average number of citations.
TNCS	Impact	Total normalized number of citations.
MNCS	Impact	Average normalized number of citations.
P(top n%)	Impact	Total number of publications that belong to the top n% of their field.
PP(top n%)	Impact	Proportion of publications that belong to the top n% of their field.
P(uncited)	Impact	Total number of uncited publications.
PP(uncited)	Impact	Proportion of uncited publications.
% Self cit	Impact	Proportion of self-citations.
TNJS	Journal impact	Total normalized journal score
MNJS	Journal impact	Mean normalized journal score, that is, the average MNCS value of the journals in which the publications of a research unit have appeared.
Int cov	Output	Internal coverage. Measured by the proportion of cited references in the analysed publications of a research unit that point to publications indexed in WoS.

3. Overall results

In this chapter, the results of the performance analysis of the Centres of Excellence (CoEs) are reported at the aggregate level. The results discussed in this chapter are based on the fractional counting method. Section 3.1 presents the internal coverage while Section 3.2 shows the overall results.

3.1. Coverage of publications

This study is limited to publications included in the WoS database which also meet the requirements as set out in Chapter 2. To what extent are the references in the collected set of publications referring to other publications which are included in the WoS? Answering this question will give insight into the WoS internal coverage of the publications included in this study. To this end, references in the relevant CoE publications (2006-2013) were matched with our WoS publication database (1980-2013). This gives an impression of the extent to which the covered publications themselves cite WoS papers, and to what extent they cite other non-WoS documents. The overall results (see Table 3.1) show that the internal coverage of the publications considered is around 85% for both periods of analysis (2006-2009 and 2010-2013). For the internal coverage, only references to core publications since 2000 are considered. Thus the internal coverage indicates the references referring to core publications published since 2000 as share of all references referring to publications published since 2000. 85% is a high internal coverage, which indicates that a bibliometric analysis for the CoEs as a whole is suitable and informative.

3.2. Publication output and citation impact by CoEs

In this section, publication output and citation impact results are presented for the aggregate of CoEs. The first subsection focuses on the CoEs at the aggregate level, while the second subsection describes the performance for the CoEs at the level of the five main fields as defined in the CWTS Leiden Ranking.

3.2.1. CoEs: performance at the aggregate level

The indicators on publication output and citation impact are presented in Table 3.1. For the citation indicators MNCS, PP(top 10%), and MNJS, Figures 3.3, 3.4, and 3.5 are informative.

Table 3.1 Bibliometric indicators for Academy of Finland's Centres of Excellence (06-09/10 and 10-13/14, fractional counting)

Unit of Analysis	P	TCS	P (top 10%)	MCS	MNCS	MNJS	PP (top 1%)	PP (top 10%)	PP (top 50%)	PP (uncited)	% Self cits	Int Cov
CoEs 2006-2009/2010	2,288.1	13,338.3	309.9	5.8	1.28	1.25	1.49%	13.5%	59%	23%	27%	85%
CoEs 2010-2013/2014	2,401.6	20,224.0	290.0	8.4	1.20	1.18	1.23%	12.1%	60%	13%	27%	85%

Table 3.2 Bibliometric indicators for Academy of Finland's Centres of Excellence (2006-2013/14, full counting)

Unit of Analysis	P	MNCS	PP (top 1%)	PP (top 10%)	Int cov
CoEs (all addr)	7,099	1.53	2.32%	15.9%	86%
CoEs (only Finland addr)	6,836	1.44	2.06%	15.4%	86%

The total fractionalized output by the CoEs between 2006 and 2009 is 2,288.1 publications and has increased during the period 2010-2013 to 2,401.6 publications. Accounting for differences in fields and publication year, the citation impact is above world average. Although the citation impact increases in absolute terms (MCS from 5.8 to 8.4), the relative citation impact (MNCS) decreases. The MNCS drops from 1.28 for 2006-2009 to 1.20 for 2010-2013. Next to the MNCS, highly cited publications are another dimension of citation impact. Compared with the mean based indicators, this approach is less susceptible to influences by outliers. The share of highly cited publications is above the expected values of 1% (PP(top 1%), 10% (PP(top 10%)) and 50% (PP(top 50%)). During the recent period, the share of highly cited publications decreases as compared with the period 2006-2009.

The journals which have been chosen for publication by CoEs have an impact value above the world average (2006-2009: MNJS = 1.25; 2010-2013: MNJS = 1.18). As with the other citation impact indicators, these scores indicate a slight decrease in citation impact. Moreover, the scores show that on average the journal citation impact is comparable to the citation impact of CoE publications.

The share of uncited publications has dropped over the years (from 23% to 13%). Self-citations account for about a quarter of all citations during both periods of analysis. A self-citation occurs if a citing publication contains the same surname and initial(s) for at least one author as in the cited publication. During both periods of analysis, almost two thirds of all publications have some degree of collaboration, i.e. about 65% of the publications contain at least two different addresses. Scientific collaboration by the CoEs is discussed in more detail in Chapter 4.

To summarize these results, the CoEs' citation impact in terms of MNCS and the share of highly cited publications has decreased. The journal citation impact is comparable to the publication citation impact. On indicators which do not measure citation impact, the CoEs have shown some development (decrease of uncitedness, increase of (international) collaboration) or stability (share of self-citations).

Exclusion of publications without Finnish address

Since not all publications by CoEs have been included in the analysis based on fractional counting, the group of publications included in the full counting analysis has been split in two: a group including publications with at least one Finnish address and a group which includes all CoE-publications. Table 3.2 presents the relevant citation indicators (note the period of analysis: 2006-2013/2014 which creates a longer citation window for the publication years 2006 to 2009). The internal coverage is similar for both publications sets. The citation impact of the group which includes all publications is higher than the citation impact of the publications included in the citation impact analysis based on fractional counting. It is impossible to estimate reliably how the citation impact would be influenced if those publications had been included, because this depends on the number of Finnish addresses as share of all addresses on those publications. The effect of inclusion would not be as strong as presented in Table 3.2 because of the differences between fractional counting and full counting. Since fractional counting only partially assigns the citation impact to the Centres of Excellence as compared with full assignment in case of full counting. This yields a higher citation impact in case of full counting, because particularly publications resulting from international collaboration – the publications which are only partially taken into account in case of fractional counting – usually have a higher citation impact than other publications (see also Chapter 4).

3.2.2. CoEs: five main fields

This analysis also pays attention to the five main fields of activity as defined by the CWTS Leiden Ranking. The five main fields are: Biomedical and health sciences, Life and earth sciences, Mathematics and computer science, Physical sciences and engineering, Social sciences and humanities. Although publications only belong to one subfield, publications can belong to multiple main fields. Such publications are assigned to a field by a weight which equals one divided by the number of assigned fields.

Output

Tables 3.3 and 3.4 give an overview of the bibliometric indicators for both periods of analysis at the level of the five main fields. To support the comparison between the two periods at the level of fields, the most important indicators are presented in Figures 3.1 to 3.4.

Two fields stand out for their share in the CoEs research output (Figure 3.1): Biomedical and health sciences, and Physical sciences and engineering. During the period of analysis, Life and earth sciences demonstrate a biggest increase in terms of publication output. The research output in Mathematics and computer science, and

Social sciences and humanities is limited in WoS core journals. At the same time, the internal coverage of the former fields (> 0.8) is also higher than latter fields (0.5 – 0.7). This might indicate that, although the research output in Mathematics and computer science, and Social sciences and humanities is limited in this analysis, a substantial part of the research output may be uncovered by the WoS database. Therefore, the results in those fields should be interpreted more carefully. Comparing the development of the research output, the share of Life and earth sciences increases substantially and the share of publications in Physical sciences and engineering decreases. The share of output in the other three fields remains the same (Figure 3.2).

Impact

As already mentioned, the citation impact of the CoEs at the aggregate level decreases during the period of analysis. Figures 3.3 and 3.4 indicate that the citation impact clearly decreases in Life and earth sciences and Physical sciences and engineering. The citation impact increases or remains equal in Biomedical and health sciences. In Mathematics and computer science and Social sciences and humanities, it is not clear if the actual citation impact increases or decreases, because the performance in terms of MNCS and PP(top 10%) shows an inconsistent development. The share of highly cited publications increases in Mathematics and computer science while it decreases in Social sciences and humanities. These are also the fields which have a lower internal coverage. The internal coverage does not indicate that a bibliometric analysis is not useful, but the citation impact indicators for these fields should be interpreted more carefully, since it is conceivable that important scientific contributions have been made which have not been indexed in the Web of Science.

In the majority of fields, the citation impact is substantially above world average, even if the citation impact decreases in some fields. There is only one conspicuous exemption: the citation impact of publications in Physical sciences and engineering drops from considerably above world average (MNCS) or the expected value (PP(top 10%) to around world average or below the expected value.

Comparing the MNCS and MNJS gives an impression of the average journal citation score of the journals selected for publication as compared with the citation performance of the unit of analysis. Although the score of the CoEs on both citation impact indicators is more or less equal during the period of analysis, there is quite some diversion at the level of main fields. The MNCS and MNJS are similar during both periods of analysis in the main fields Biomedical and health sciences, and Life and earth sciences. Although the MNCS drops in Mathematics and computer science, the MNCS is still considerably higher than the MNJS, which remains stable over the years. Thus, on average, publications in Mathematics and computer science attract more citations than all publications in the journals in these fields in which CoE-publications have been published. On the contrary, the drop in MNCS in the field Physical sciences and engineering goes hand in hand with a drop in MNJS. This is not the case for the Social sciences and humanities, in which the MNCS increases while the MNJS remains stable.

Other indicators

Tables 3.3 and 3.4 also present some indicators which do not measure publication output or citation impact. The share of self citations ranges from 22% to 33%. The uncitedness is high in fields with a low citation density as indicated by the mean citation score (MCS) (Mathematics and computer science, and Social sciences and humanities). An increase in MCS correlates with a decrease in uncitedness. The other fields show a decrease in uncitedness as well, and particularly Physical sciences and engineering develop more towards the other two fields (Biomedical and health sciences, and Life and earth sciences) with a low level of uncitedness.

Exclusion of publications without Finnish address

In Table 3.5 the publications which have at least one Finnish address and which have been included in the fractional counting analysis are compared at the level of main fields with the publication set which includes both the publications included in the citation impact analysis and the publications excluded from it (i.e. publications without a Finnish address). The publication set which includes all addresses has a higher citation impact than the publications which have at least one Finnish address. This is particularly the case in the field Biomedical and health sciences. However, differences are mainly obvious for the MNCS indicator and the PP(top 1%) indicator. The citation impact in terms of the PP(top 10%) indicator is more or less equal for both publication sets. Since the PP(top 1%) and PP(top 10%) have a strong correlation with the MNCS, the MNCS is not heavily affected by extreme outliers. The internal coverage is similar in all fields. As mentioned before, the citation impact is higher when publications would have been included that have now been excluded from the citation impact analysis in this report because of missing Finnish addresses. However, it is unclear to what extent this would influence the results of the analysis based on fractional counting, but the difference between the two sets is probably smaller if both publication sets could have been included in citation analysis based on fractional counting.

Table 3.3 Bibliometric indicators for Academy of Finland's Centers of Excellence: five main fields (06-09/10, fractional counting)

Main field	P	TCS	P (top 10%)	MCS	MNCS	MNJS	PP (top 1%)	PP (top 10%)	PP (top 50%)	PP (uncited)	% Self cits	Int Cov
Biomedical and health sciences	786.1	6,038.9	99.8	7.68	1.23	1.29	1.46%	12.7%	60%	15.1%	22.1%	92%
Life and earth sciences	469.3	2,780.0	67.9	5.92	1.31	1.33	1.42%	14.5%	62%	16.1%	29.3%	85%
Mathematics and computer science	208.8	556.6	29.6	2.67	1.37	1.12	1.39%	14.2%	59%	41.3%	25.3%	52%
Physical sciences and engineering	684.7	3,351.9	94.6	4.90	1.32	1.21	1.85%	13.8%	57%	26.7%	32.7%	86%
Social sciences and humanities	139.2	610.8	18.0	4.39	1.05	1.13	0.21%	12.9%	58%	37.3%	27.3%	69%

Table 3.4 Bibliometric indicators for Academy of Finland's Centers of Excellence: five main fields (10-13/14, fractional counting)

Main field	P	TCS	P (top 10%)	MCS	MNCS	MNJS	PP (top 1%)	PP (top 10%)	PP (top 50%)	PP (uncited)	% Self cits	Int Cov
Biomedical and health sciences	819.8	8,692.3	110.2	10.60	1.27	1.28	1.42%	13.4%	62%	7.3%	22.5%	91%
Life and earth sciences	573.8	4,410.3	71.3	7.69	1.18	1.21	0.84%	12.4%	60%	11.7%	32.8%	86%
Mathematics and computer science	205.4	855.2	30.2	4.16	1.27	1.12	0.71%	14.7%	61%	30.8%	22.4%	55%
Physical sciences and engineering	663.6	5,533.2	61.0	8.34	1.09	1.07	1.46%	9.2%	55%	15.5%	29.0%	86%
Social sciences and humanities	139.0	733.0	17.2	5.27	1.25	1.11	1.43%	12.4%	62%	19.1%	28.4%	67%

Table 3.5 Bibliometric indicators for Academy of Finland's Centers of Excellence: five main fields (2006-2013/14, full counting)

	Publications with Finnish addresses						Publications with and without Finnish addresses				
Main field	**P**	**MNCS**	**PP (top 1%)**	**PP (top 10%)**	**Int cov**		**P**	**MNCS**	**PP (top 1%)**	**PP (top 10%)**	**Int cov**
Biomedical and health sciences	2,291.7	1.59	2.85%	17.3%	92%		2,395.0	1.80	3.46%	18.4%	91%
Life and earth sciences	1,578.5	1.47	1.93%	15.5%	86%		1,646.5	1.47	1.91%	15.7%	86%
Mathematics and computer science	568.2	1.45	1.56%	17.3%	55%		592.5	1.56	1.84%	17.8%	56%
Physical sciences and engineering	2,040.7	1.28	1.69%	12.9%	86%		2,094.0	1.29	1.69%	13.2%	86%
Social sciences and humanities	357.0	1.25	0.87%	13.5%	71%		372.0	1.27	1.10%	13.4%	70%

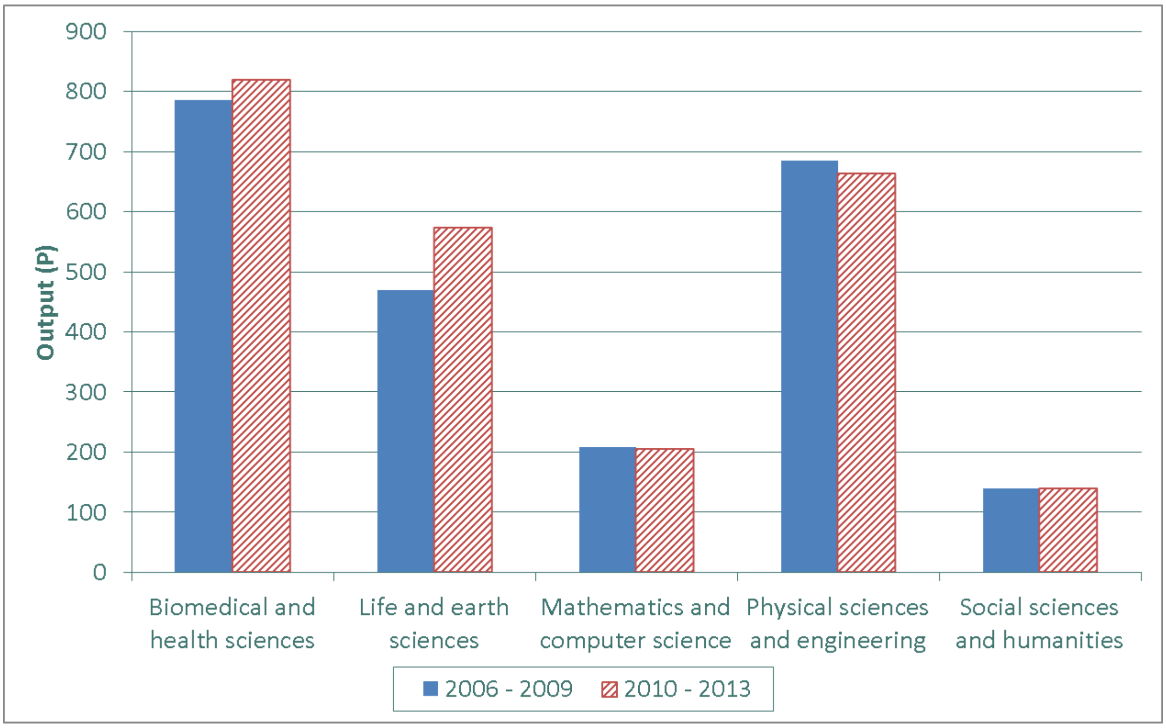

Figure 3.1. Main fields of activity for CoEs: publication output (absolute)

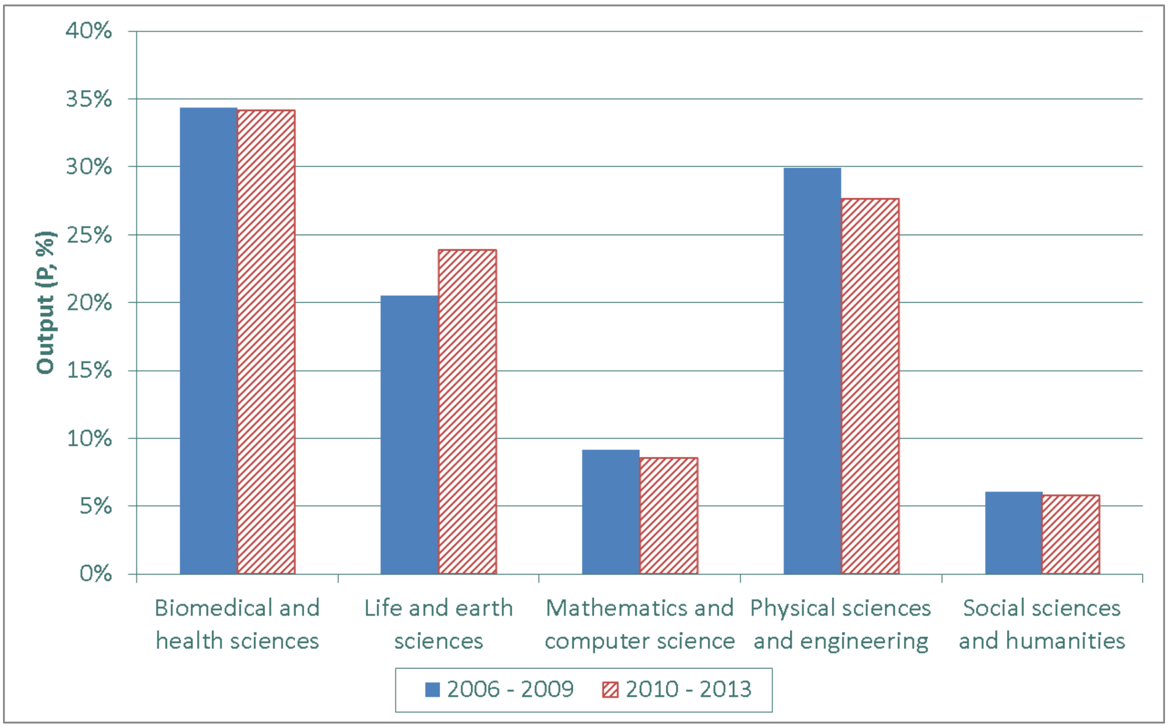

Figure 3.2. Main fields of activity for CoEs: publication output (relative)

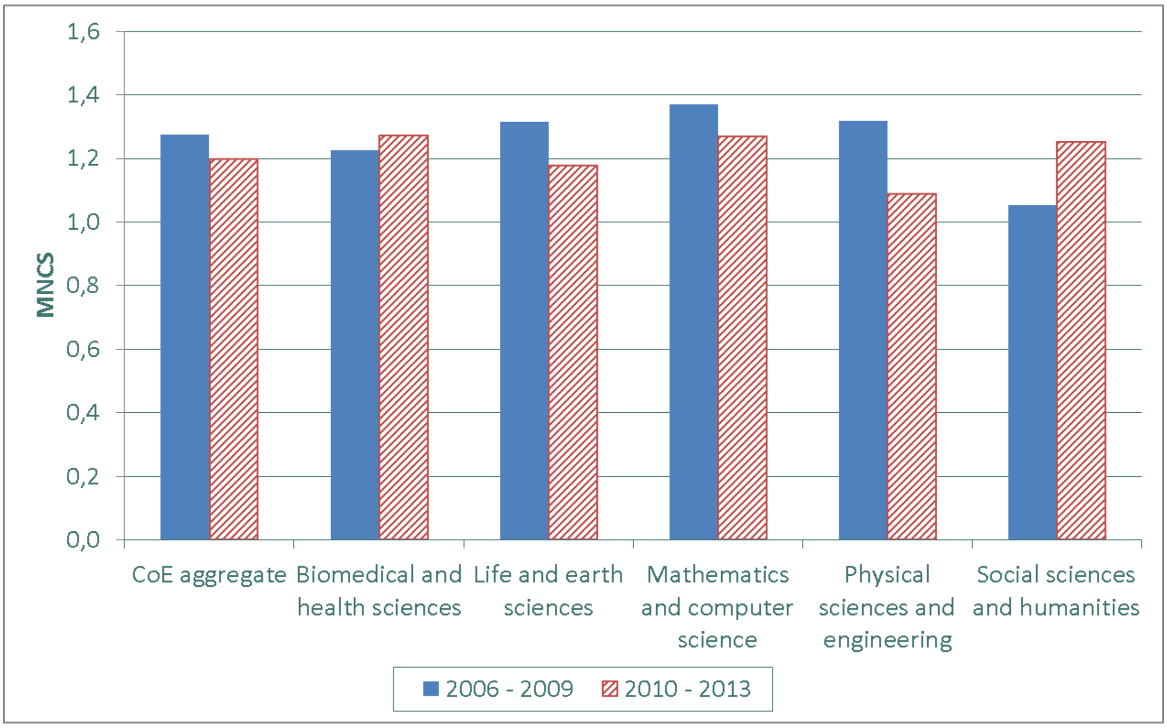

Figure 3.3. Main fields of activity for CoEs: citation impact (MNCS)

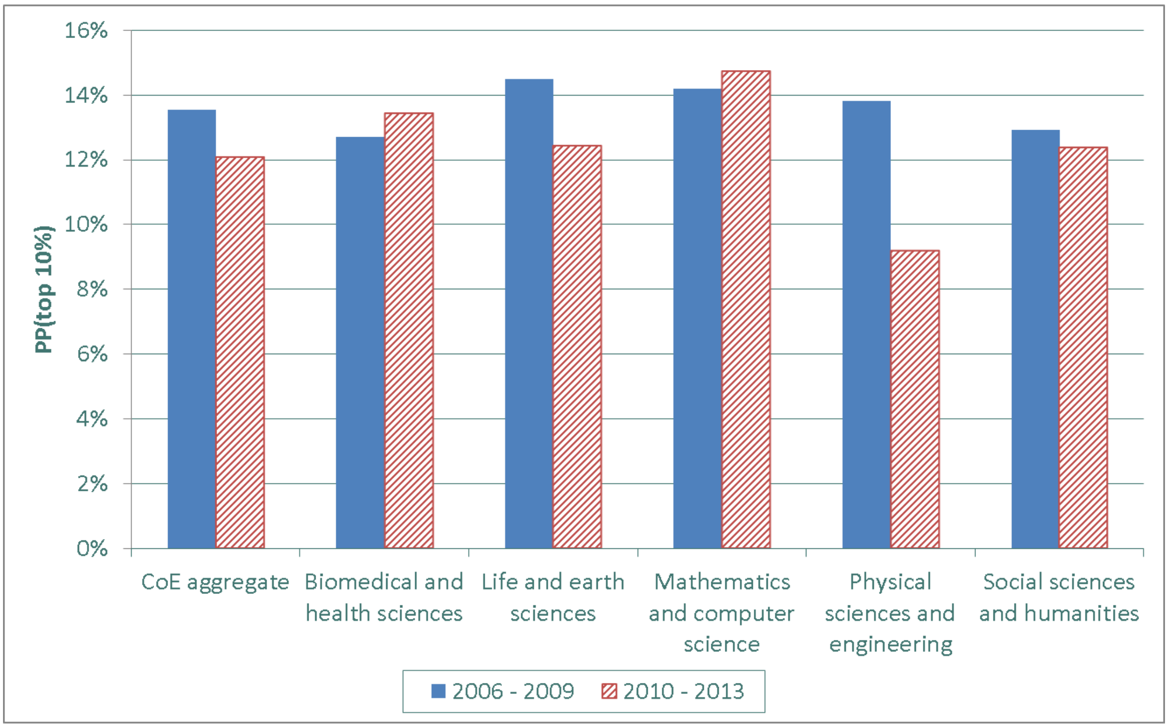

Figure 3.4. Main fields of activity for CoEs: citation impact (PP(top 10%))

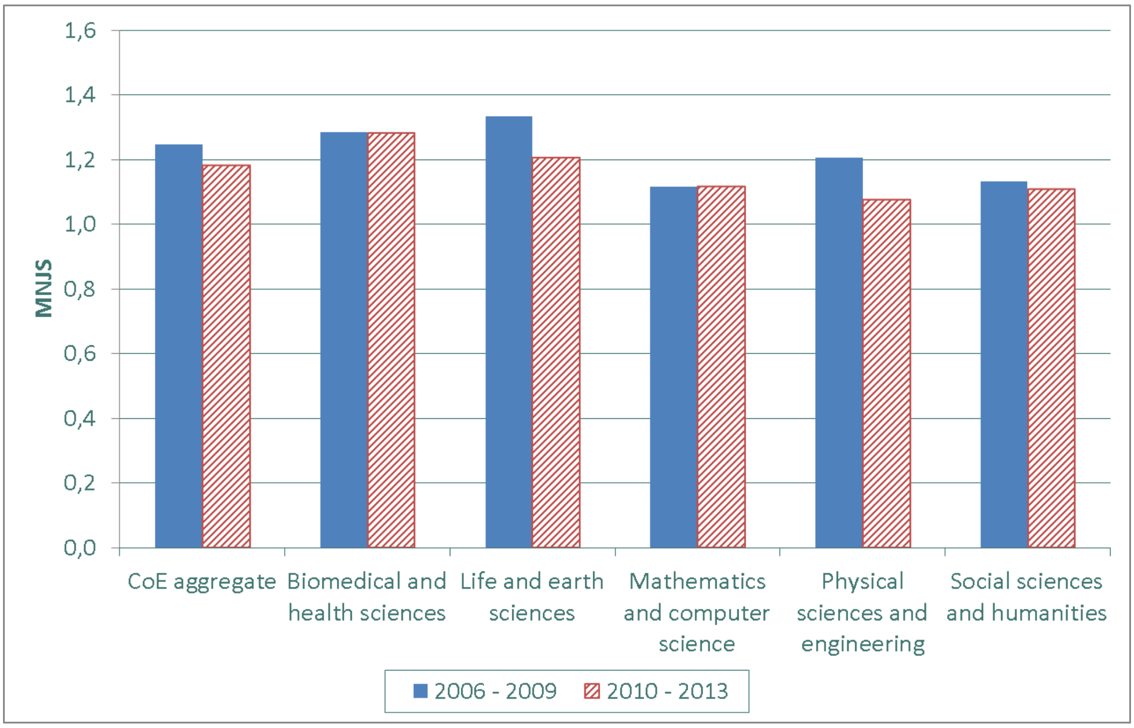

Figure 3.5. Main fields of activity for CoEs: citation impact (MNJS)

3.3. Impact by Centres of Excellence on Finnish science

Table 3.6 presents the main publication and citation indicators for Finland during the periods 2006-2009 and 2010-2013 both including and excluding the publications by the CoEs. This analysis is only relevant for the CoE publications which have been included in the fractional counting analysis, i.e., the CoE publications which contain at least one Finnish address.

The figures in Table 3.6 reveal that the publication set which includes the publications by CoEs have a higher citation impact regarding any of the presented citation impact indicators. There is a quite substantial difference in citation impact between the two complementary publication sets: publications by the CoEs (Table 3.1) and all publications which mention Finland as (one of) their affiliation(s) excluding the CoEs (right half of Table 3.6). Nevertheless, the higher citation impact for publications by CoEs does not result in a much higher citation impact at the level of Finland as a whole (see Table 3.6). This is caused by the relatively small set of publications by CoEs as compared with all publication with at least one Finnish affiliation during the periods of analysis.

Table 3.6 Bibliometric indicators for Finland including CoEs and excluding CoEs (2006-2013/14, fractional counting)

	Finland including CoEs						Finland excluding CoEs				
Main field	P	MNCS	PP (top 1%)	PP (top 10%)	Int cov		P	MNCS	PP (top 1%)	PP (top 10%)	Int cov
2006-2009/2010	25,407.9	0.97	0.83%	9.4%	78%		23,119.9	0.94	0.76%	9.0%	78%
2010-2013/2014	27,418.0	1.00	0.91%	9.6%	78%		25,016.4	0.98	0.88%	9.4%	77%

4. Scientific collaboration

In this chapter the main collaborative patterns are presented. Next to the collaboration analysis at the level of international, national and non-collaborative papers (section 4.2), we discuss collaboration with industry (section 4.3). The output statistics are presented both in terms of full counting and fractional counting, all citation indicators are based on fractional counting.

4.1. Types of collaboration

The different collaboration types are identified on the basis of address information in the WoS address field:

- No collaboration: the address information in the publication contains a single institutional address. Such publications can have multiple authors from the same institution.
- National collaboration: the address information in the publication contains addresses from one country but from at least two different institutions.
- International collaboration: the address information in the publication contains addresses from at least two countries.

In this way, three mutually exclusive definitions of collaboration are used in the analysis: ‘no collaboration’, ‘national collaboration’, and ‘international collaboration’. Note that publications by CoEs are not assigned to one or another category as such. A publication which is the result of research efforts by a single CoE may still result from national or international collaboration, depending on the institutional affiliations which the authors have included. If a publication is published by a single CoE only but includes addresses from multiple institutions which collaborate within that CoE, the publications is classified as a collaborative publication.

Table 4.1 Bibliometric indicators for Finland's Centers of Excellence: collaboration types (2006-2009/10)

Main field	P (full)	P (fractional)	TCS	P (top 10%)	MCS	MNCS	MNJS	PP (top 1%)	PP (top 10%)	PP (top 50%)	PP (uncited)	PP (self cits)	Int cov
International collaboration	1,801	831.1	6,037.3	133.7	7.3	1.41	1.41	1.54%	16.1%	62%	18%	28%	88%
National collaboration	639	639.0	3,219.0	76.7	5.0	1.11	1.18	0.79%	12.0%	56%	23%	26%	86%
No collaboration	818	818.0	4,082.0	99.5	5.0	1.27	1.14	1.97%	12.2%	56%	27%	27%	82%

Table 4.2 Bibliometric indicators for Finland's Centers of Excellence: collaboration types (2010-2013/14)

Main field	P (full)	P (fractional)	TCS	P (top 10%)	MCS	MNCS	MNJS	PP (top 1%)	PP (top 10%)	PP (top 50%)	PP (uncited)	PP (self cits)	Int cov
International collaboration	2,136	959.6	10,476.0	151.1	10.9	1.44	1.36	2.15%	15.7%	61%	12%	28%	87%
National collaboration	691	691.0	4,685.0	58.9	6.8	0.97	1.07	0.52%	8.5%	55%	12%	25%	85%
No collaboration	751	751.0	5,063.0	79.9	6.7	1.09	1.06	0.72%	10.6%	55%	17%	27%	82%

4.2. Results

Table 4.1 presents the output and impact indicators per collaboration type for 2006-2009/2010 and 2010-2013/2014. For the output statistics the full counting method gives the most reliable impression of the importance of the different collaboration types, because fractional counting corrects for collaboration. The citation impact indicators presented in this chapter have been calculated on the basis of fractional counting. Output and citation indicators are also graphically presented in Figures 4.1 to 4.5 .

Centres of Excellence

International collaboration is by far the most important collaboration type for the CoEs. In absolute terms, there is an increase of international collaboration and national collaboration, but a decrease of non-collaborative publications (Figure 4.1). During the period of analysis, the share of international collaboration increases from around 55% to 60%. This increase in publications stemming from international collaboration goes hand in hand with a decrease of non-collaborative publications, while the share of publications resulting from national collaboration does not change (Figure 4.2). The internal coverage is similar for all collaboration types (Table 4.1).

Publications resulting from international collaboration have the highest citation impact in terms of MNCS, PP(top 1%), PP(top 10%), and PP(top 50%) (Figures 4.3 and 4.4 and Table 4.1). During both periods of analysis, the citation impact of publications resulting from international collaboration is around 1.4 (MNCS), the share of highly cited papers is around 16.0%, and the citation score of the journals selected for publication is around 1.4 (MNJS). The citation impact for the other types of collaboration (national collaboration, no collaboration) shows a (substantial) decrease in terms of MNCS, MNJS, and PP(top 10%). In case of non-collaborative publications, the citation impact decreases from substantially above world average to just above world average.

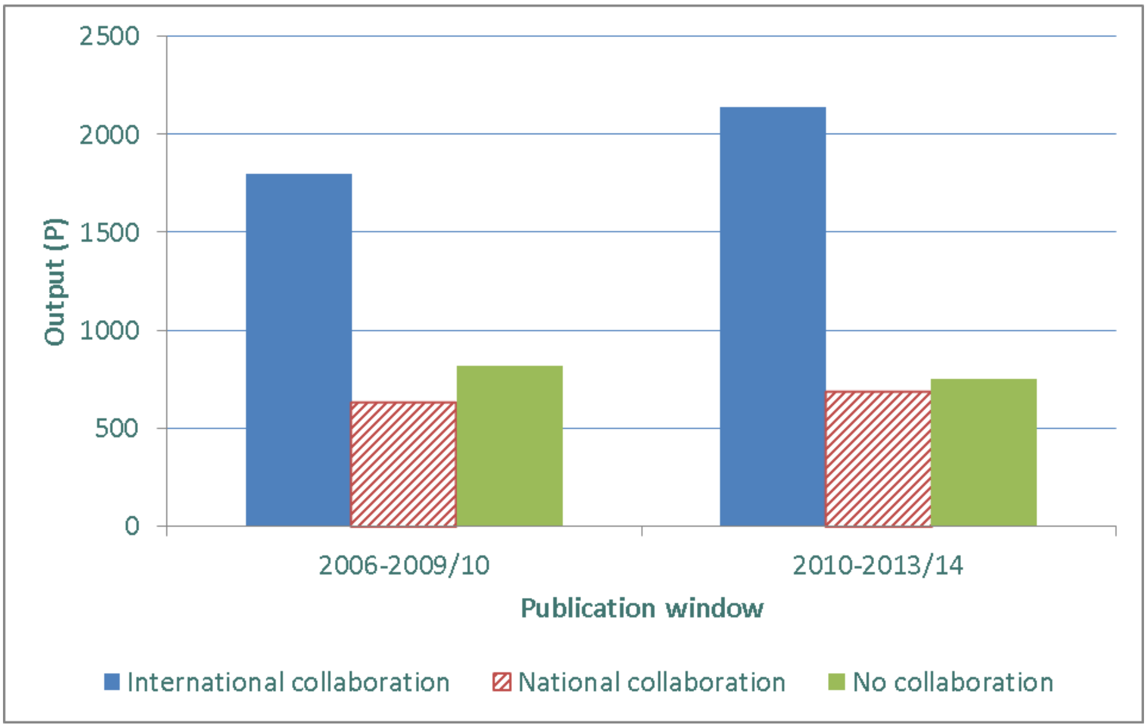

Figure 4.1. Collaboration profile of CoEs: publication output (absolute, full counting)

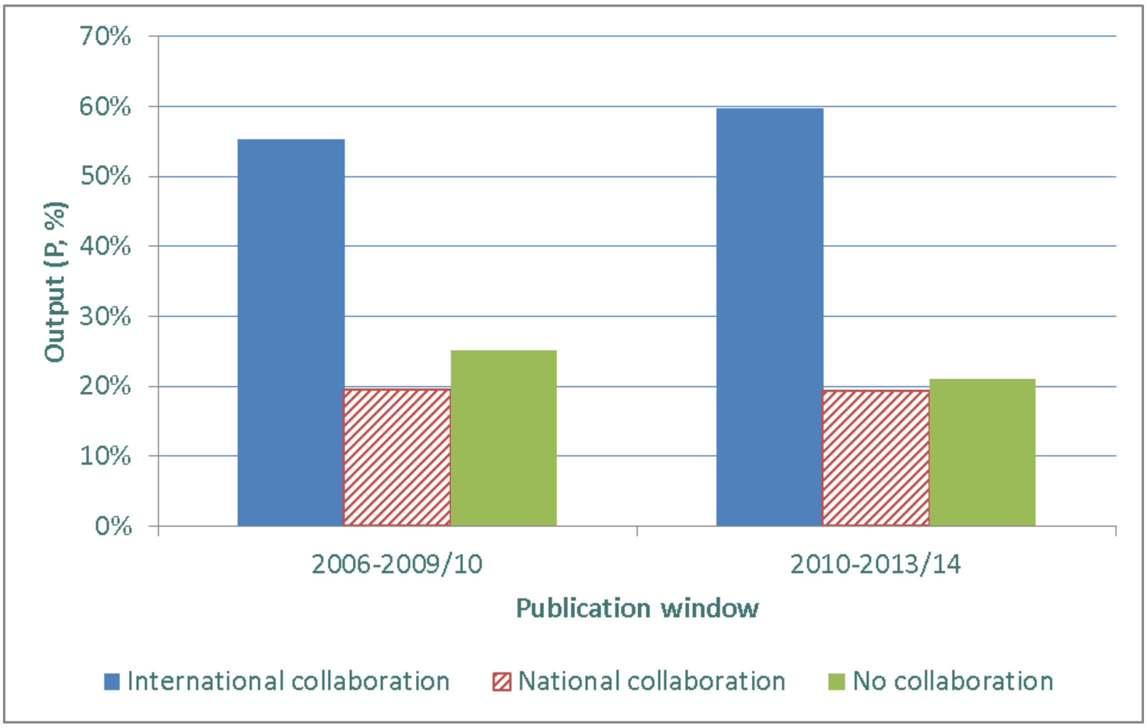

Figure 4.2. Collaboration profile of CoEs: publication output (relative, full counting)

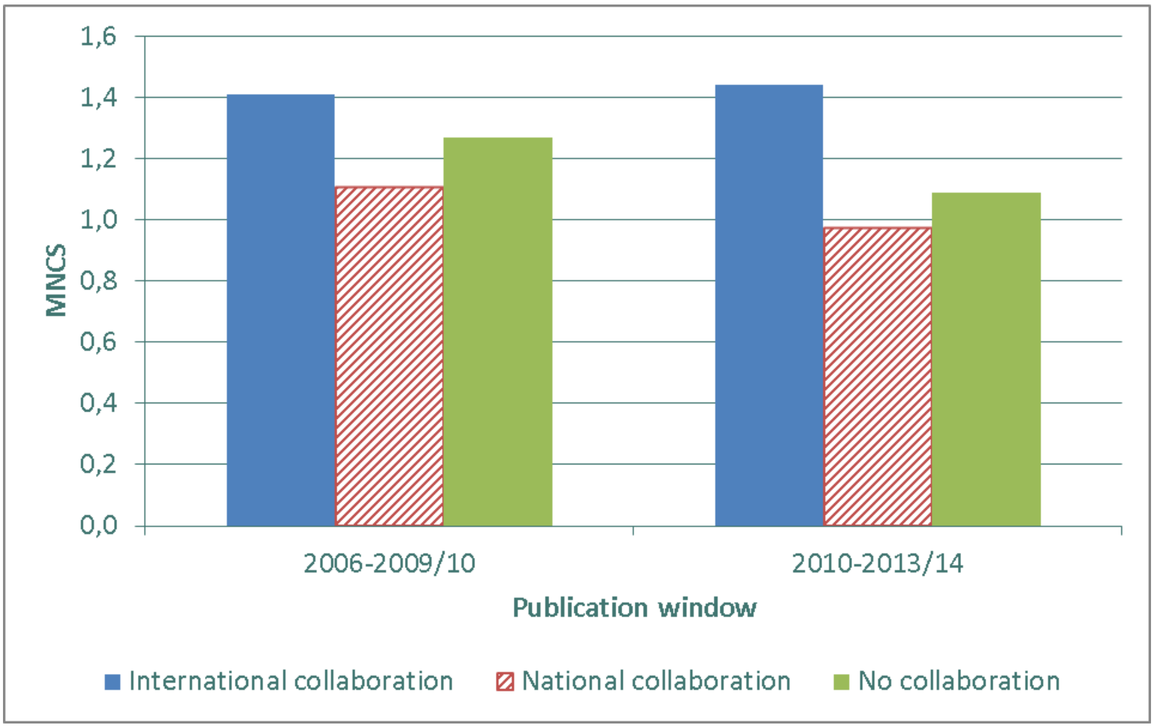

Figure 4.3. Collaboration profile of CoEs: citation impact (MNCS)

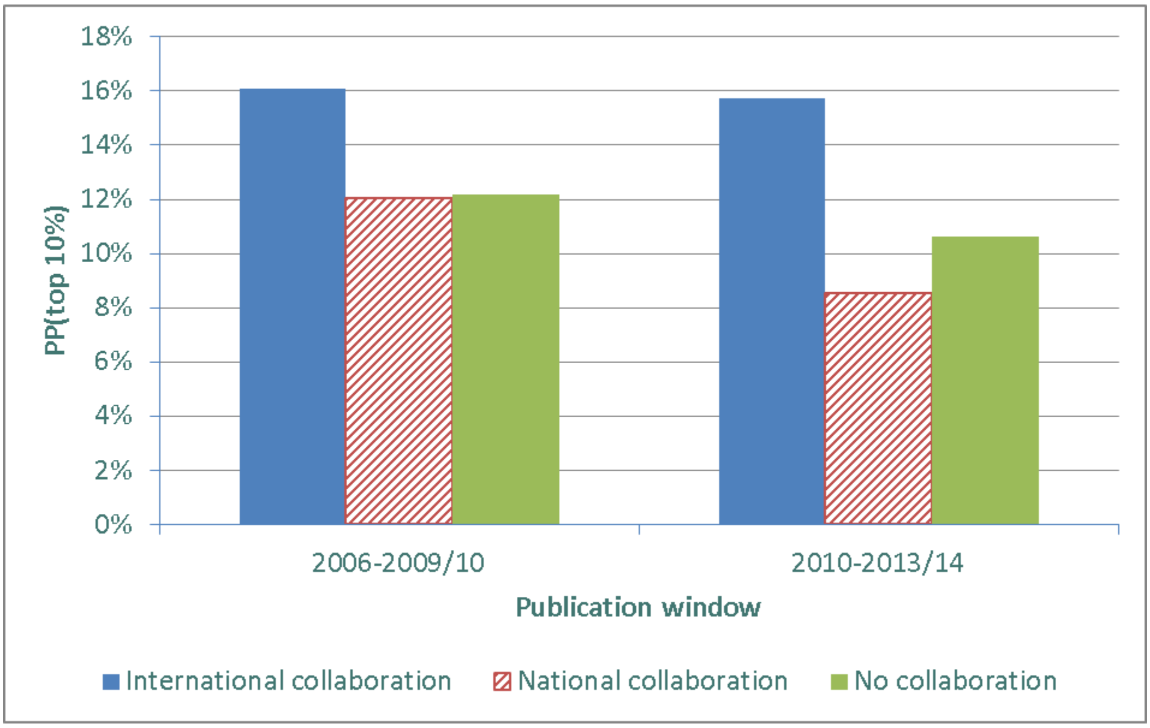

Figure 4.4. Collaboration profile of CoEs: citation impact (PP(top 10%))

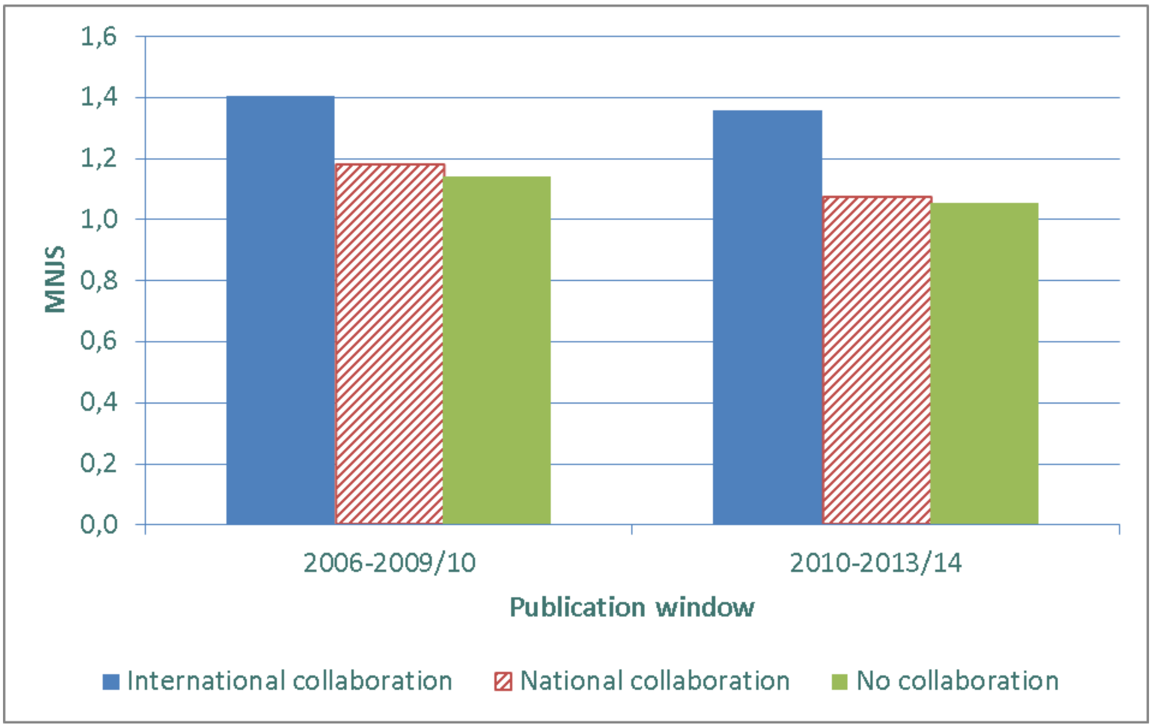

Figure 4.5. Collaboration profile of CoEs: citation impact (MNJS)

Five main fields

The collaboration profiles per main field of activity are presented in Figures 4.6 to 4.10. International collaboration is by far the most important collaboration in Biomedical and health sciences, Life and earth sciences, and Physical sciences and engineering. In Mathematics and computer science, and Social sciences and humanities, non-collaborative publications have similar importance as internationally collaborative publications during the period 2006-2009. Except for Social sciences and humanities, more than half of the research output results from international collaboration during the most recent period of analysis (2010-2013). In all fields except Biomedical and health sciences the importance of non-collaborative publications has decreased in 2010-2013 as compared with 2006-2009 (see Figures 4.6 and 4.7).

For the majority of fields, the publications resulting from international collaboration have the highest citation impact in terms of MNCS, PP(top 10%), and MNJS (see Figures 4.8 to 4.10). The citation impact on any of these indicators is very stable during both periods of analysis for the three different collaboration types for the field Biomedicine and Health sciences. The citation impact of publications resulting from international collaboration is above world average and rather stable in all fields, but the differences are particularly found in publications with another collaboration type. In Life sciences and environmental sciences, and in Mathematics and computer science, there is a considerable decrease in citation impact of publications resulting from national collaboration. This is also the case for non-collaborative publications in the field Physical sciences and engineering. The citation impact for non-collaborative publications increases considerably in the field of Social sciences and humanities. In most fields, MNJS remains rather stable for internationally collaborative publications, but there is often a slight decrease in normalized journal citation impact in the journals selected for publications not resulting from any collaboration or resulting from national collaboration (Figure 4.10). The only exceptions are Biomedical and health sciences (2006-2009 and 2010-2013) and Mathematics and computer science (non-collaborative publications, 2010-2013).

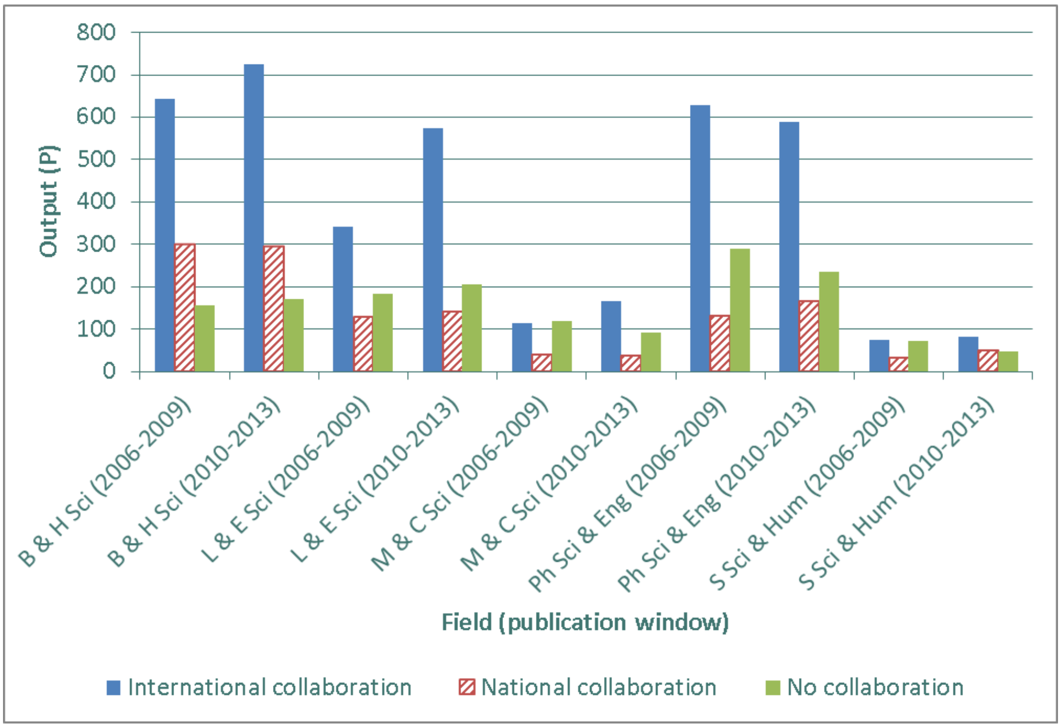

Figure 4.6. Collaboration profile per main field: publication output (P, full counting)

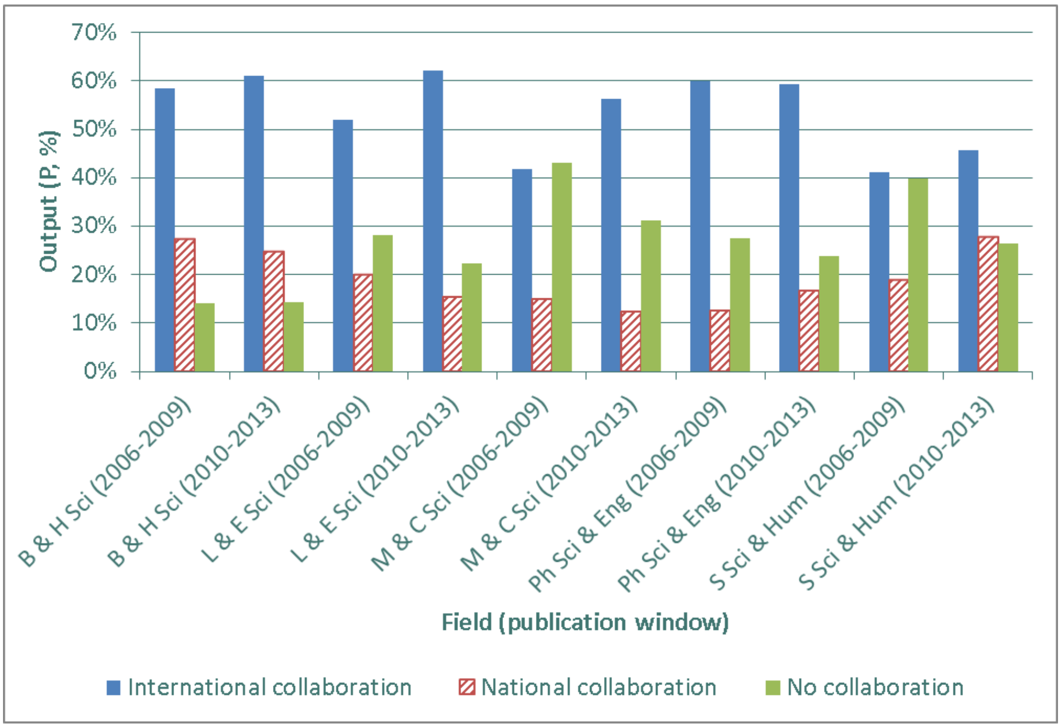

Figure 4.7. Collaboration profile per main field: publication output (P, %, full counting)

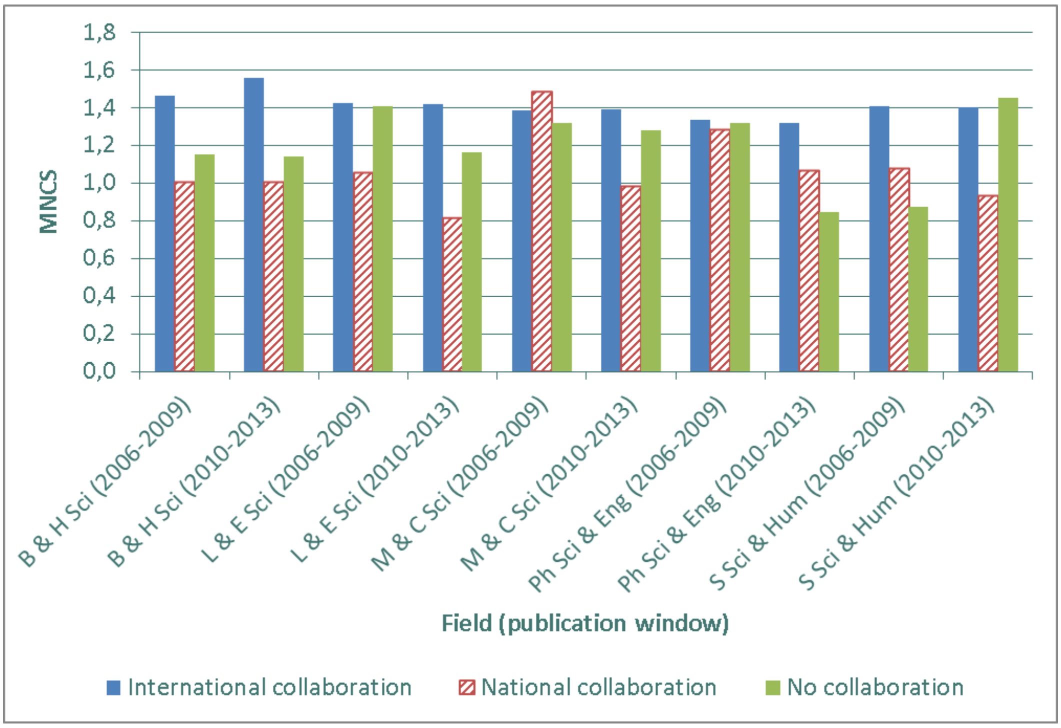

Figure 4.8. Collaboration profile per main field: citation impact (MNCS)

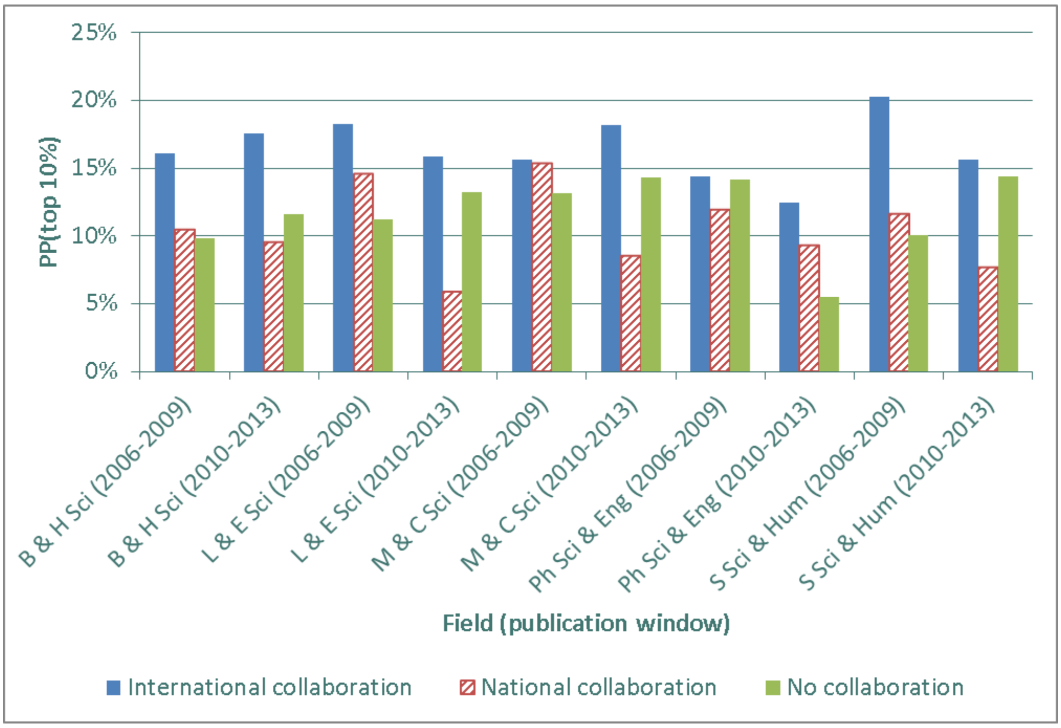

Figure 4.9. Collaboration profile per main field: citation impact (PP(top 10%))

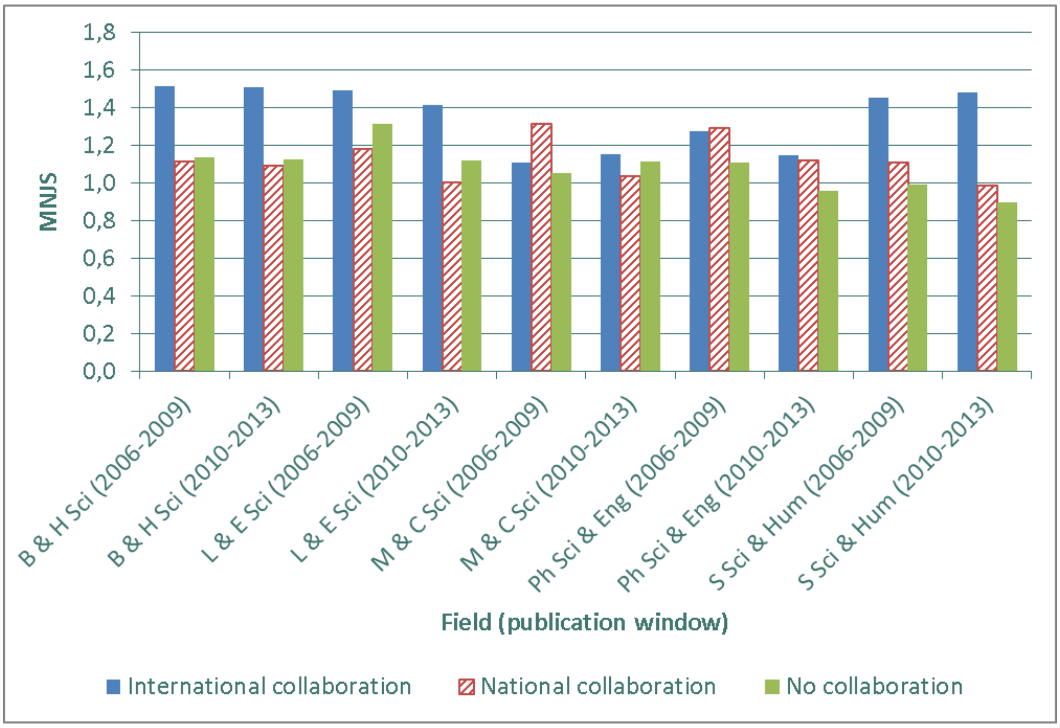

Figure 4.10. Collaboration profile per main field: citation impact (MNJS)

Note that the citation impact indicators in the fields Mathematics and computer science, and Social science and humanities are based on low output numbers. Therefore, the citation impact indicators may have a lower stability (Figures 4.7 to 4.10).

4.3. Industrial collaboration

Industrial collaboration takes place if companies are involved in a collaborative publication with a non-industrial partner such as a research institution or a university. Thus, a publication which results from industrial collaboration is never a publication which belongs to the category ‘no collaboration’. However, the distinction between national and international collaboration is not made, so a publication resulting from industrial collaboration may either be classified as ‘national collaboration’ or ‘international collaboration’. To decide if a publication results from industrial collaboration, the addresses included in the publication are analyzed. Although the ideal is to provide 100% accuracy, this cannot be reached, because publication addresses are manually processed for millions of records.

Aggregate level

The results of the industrial collaboration analysis are included in Tables 4.3 and 4.4 (the aggregate of all CoEs) and in Figures 4.11 to 4.16 (aggregate of all CoEs by main field). Only a small share of the publications by CoEs are published in collaboration with an industrial partner. By far the majority of publications have no industrial involvement. The internal coverage is similar for both sets of publications during both periods of analysis.

During the period 2006-2009, the citation impact indicators do not consistently make clear which publication set has a higher citation impact. The publication set which includes all industrial collaborative publications has a higher MNCS but a lower PP(top 10%) than the publication set excluding such publications. For both sets, the MNCS and MNJS are more or less equal. Comparing the period 2010-2013 with the period 2006-2009, the industrial collaborative publications have increased stronger (in relative terms) than the publications not resulting from industrial collaboration. Moreover, in the period 2010-2013, the citation impact is clearly higher for the former publication set as compared with the latter.

Table 4.3 Bibliometric indicators for Finland's Centers of Excellence: industrial collaboration (2006-2009/10)

Collaboration type	P (full)	P (fractional)	TCS	P (top 10%)	MCS	MNCS	MNJS	PP (top 1%)	PP (top 10%)	PP (top 50%)	PP (uncited)	PP (self cits)	Int cov
Industrial collaboration	179	106.6	645.0	13.1	6.1	1.53	1.53	2.32%	12.3%	60%	19%	29%	81%
No industrial collaboration	3,179	2,181.5	12,693.3	296.9	5.8	1.26	1.23	1.45%	13.6%	58%	23%	27%	85%

Table 4.4 Bibliometric indicators for Finland's Centers of Excellence: industrial collaboration (2010-2013/14)

Collaboration type	P (full)	P (fractional)	TCS	P (top 10%)	MCS	MNCS	MNJS	PP (top 1%)	PP (top 10%)	PP (top 50%)	PP (uncited)	PP (self cits)	Int cov
Industrial collaboration	316	160.3	2,002.6	26.4	12.5	1.72	1.55	3.35%	16.5%	60%	7%	32%	82%
No industrial collaboration	3,425	2241.3	18,221.4	263.5	8.1	1.16	1.16	1.08%	11.8%	58%	14%	26%	85%

Five main fields

In Figures 4.11 to 4.16, the most important output and citation impact statistics are presented at both the aggregate level and at the level of the five main fields. As with the collaboration analysis in the previous section, full counting has been applied for the analysis of the output.

In all fields industrial collaboration has become more important. In absolute terms, the highest number of publications resulting from industrial collaboration is published in the field Biomedical and health sciences. Particularly in the fields Life and earth sciences, and Physical sciences and engineering, the industrial collaborative output has increased substantially. The industrial collaboration output in the field Mathematics and computer science remains low in absolute terms. Industrial collaboration hardly plays a role in Social sciences and humanities (Figure 4.11). Since the publication output per field differs considerably, the relative output of industrial collaborative publications may be more informative (Figure 4.12). Industrial collaboration is important in Mathematics and computer science, but the other fields except Social sciences and humanities have come close during the recent period of analysis. In Social sciences and humanities, industrial collaboration is irrelevant.

The citation impact differs considerably per field. Since the citation impact is calculated on the basis of small publication sets (< 60 publications, see Figure 4.11), the results can easily be influenced by outliers and randomness. Therefore, one should not attach too much value to the citation impact analysis at the level of the main fields. Moreover, it is difficult to discern a clear pattern in the citation impact results. Publications resulting from industrial collaboration have a higher citation impact in the field of Biomedicine and health sciences, but a lower citation impact in the field of Mathematics and computer science (in which they amount to a relatively large share of the publication output). In the other fields, the results are mixed, though in the field of Physical sciences and engineering the citation impact is considerably lower during the period 2010-2013. This is also the period during which industrial collaborative publications have become relatively important in that field.

In most fields, the MNJS generally has the same pattern as the MNCS but – given the small size of the underlying publication set – may differ considerably in score.

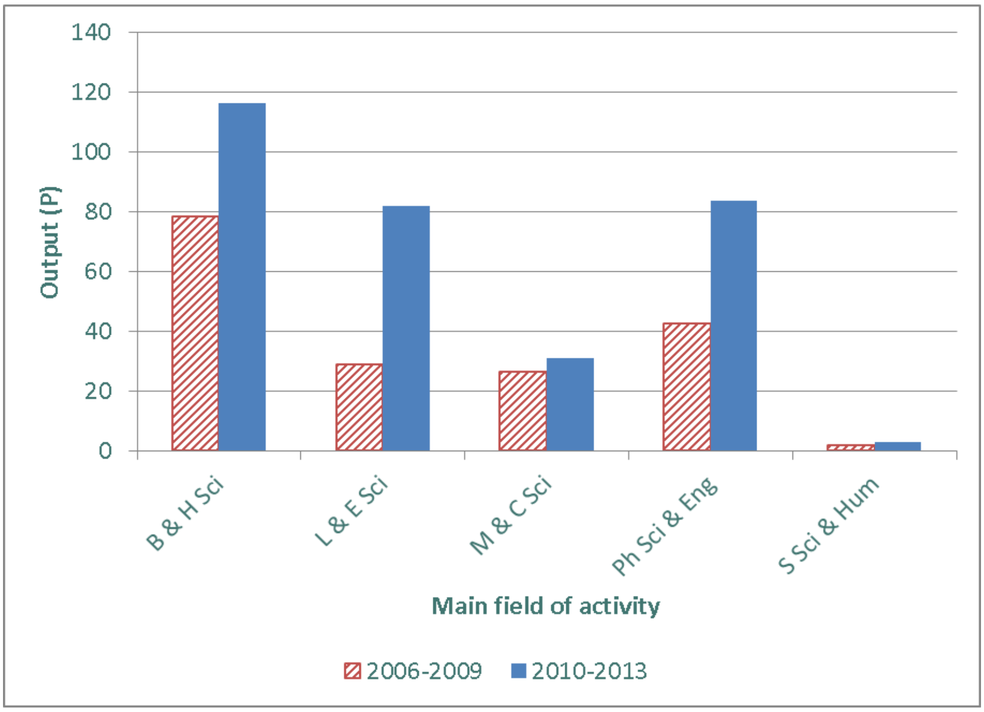

Figure 4.11. Industrial collaboration profile per main field: publication output (P, full counting)

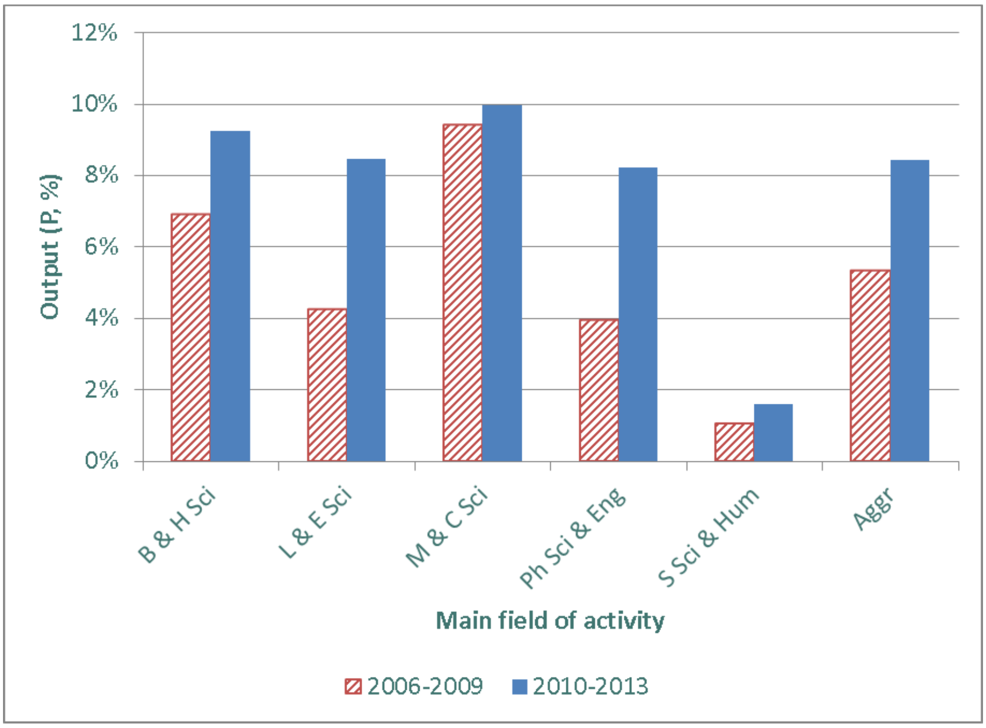

Figure 4.12. Industrial collaboration profile per main field: publication output (P, %, full counting)

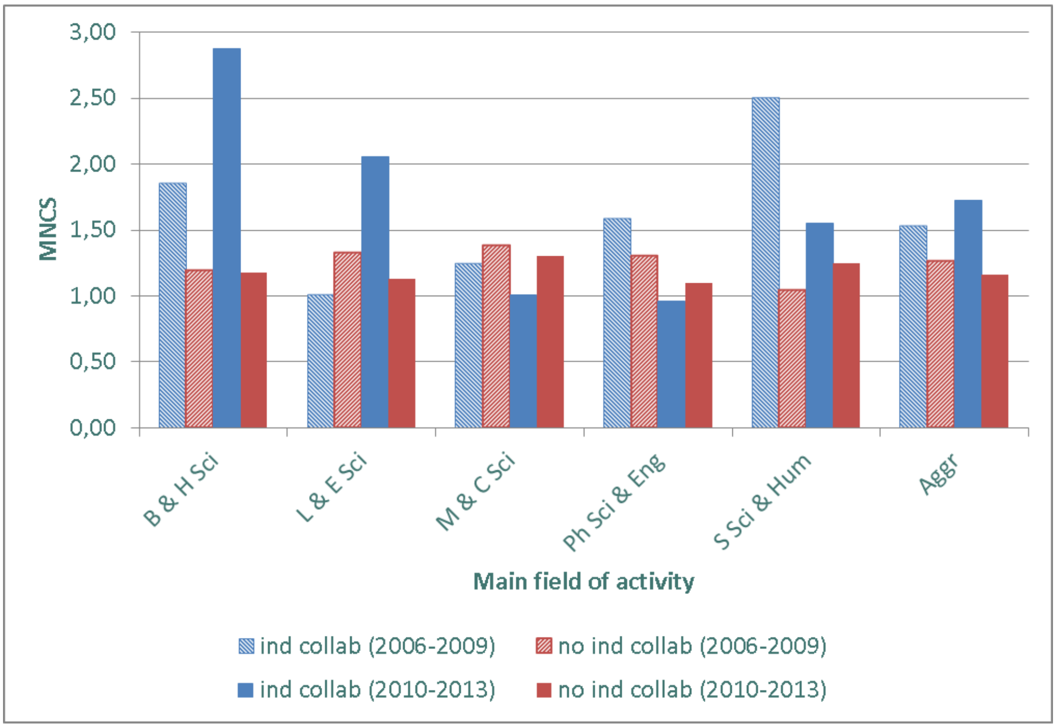

Figure 4.13. Industrial collaboration profile per main field: citation impact (MNCS)

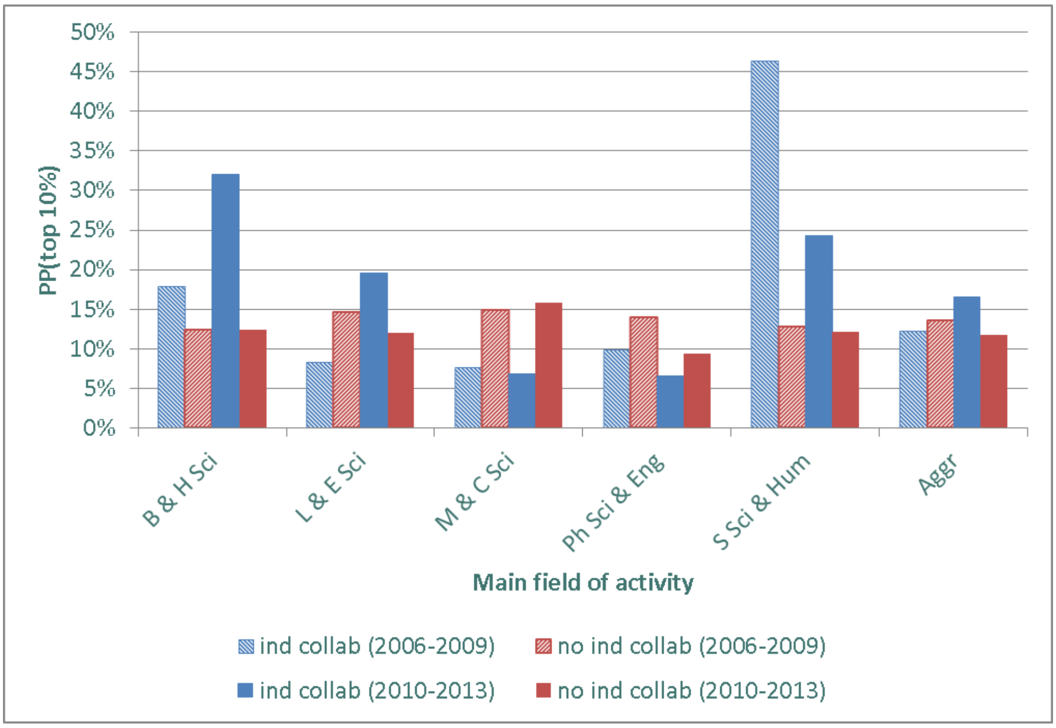

Figure 4.14. Industrial collaboration per main field: citation impact (PP(top 10%))

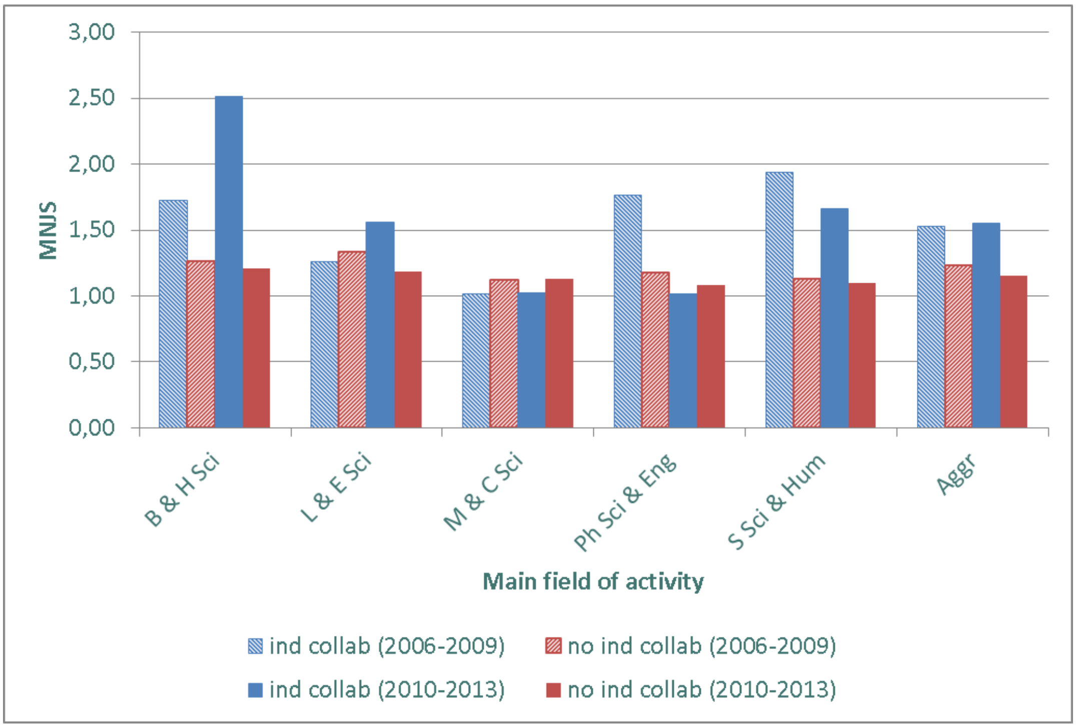

Figure 4.15. Industrial collaboration profile per main field: citation impact (MNJS)

Note that the citation impact indicators of the industrial collaboration publication sets are based on a low number of publications. This may result in a lower stability of the citation impact indicators (Figures 4.13 to 4.15).

5. Conclusions

This study has focused on the publications of Finland's Centers of Excellence during the periods 2006 to 2009 and 2010 to 2013. The analysis provides insight into the citation impact of the publications resulting from the CoEs, not only at the aggregate level, but also at the level of main fields and for different collaboration types. A fractional counting methodology has been used: publication output and citation impact have been assigned to the CoEs proportionally. This enables the comparison with the results from the CWTS Leiden Ranking.

Almost 7,000 publications have been included in the citation analysis of the Finnish Centres of Excellence. The internal coverage indicates that a bibliometric analysis is suitable for the publication output by CoEs, but the internal coverage in the fields Mathematics and computer science and Social sciences and humanities is somewhat limited. The results in these fields should be interpreted more carefully.

Over the years, there has been an increase in publication output. The citation analysis has revealed that the CoEs citation impact at the aggregate level is about 20% to 30% above world average, but has slightly decreased. The citation impact of the journals in which CoEs have published is in line with the citation impact of publications by CoEs and is above world average. Compared with the citation impact of publications affiliated with Finland, publications by the CoEs have a substantially higher citation impact, but their influence on the citation impact of all Finnish publications is limited given the vast difference in size of both publication sets.

The fields with the largest research output are Biomedical and health sciences and Physical sciences and engineering. The field of Life and earth sciences should be added to these fields on the basis of the recent publication output results. While the publication output increases considerably in the field of Life and earth sciences, there is quite a decrease in the field of Physical sciences and engineering. The fields in which the publication output is lowest (Mathematics and computer science and Social sciences and humanities) are also the fields with the lowest internal coverage, so probably those fields are the ones with the highest non-WoS publication output. The citation impact is high in most fields, but decreases substantially in Life and earth sciences and in Physical sciences and engineering.

International collaboration has become more important over the years. During the recent period of analysis (2010-2013), international publications also have become the most important type of publications in all of the five main fields (ranging from around 45% in Social sciences and humanities to more than 60% in Life and earth sciences). Internationally collaborative publications also have the highest citation impact, which remains stable over the years as well. At the aggregate level, the decrease in citation impact is caused by publications resulting from national collaboration or not resulting from any collaboration, because only the citation impact by internationally collaborative publications remains stable over the years. This pattern is, however, not present in all main fields. The importance of the collaboration types differs a lot per field.

The share of publications resulting from industrial collaboration is limited for the CoEs, but increases considerably over the years (from about 5% to over 8% of the publication output). Industrial collaboration has increased in all fields. Recently it has gained a similar relevance as it already had in Mathematics and computer science in all fields except Social sciences and humanities. In Social sciences and humanities, industrial collaboration is unimportant. At the aggregate level, the citation impact is similar or higher for the industrial collaborative publications as compared with the publications which do not result from collaboration.

The results presented in this report can be compared with the results presented in the CWTS Leiden Ranking 2015 (www.leidenranking.com). At the aggregate level, the citation impact of the CoEs during the period 2006-2009 is comparable to the citation impact by universities between ranking position 75 and 100 such as Leiden University or the Technical University of Munich. The CoEs have a higher citation impact during this period of analysis than any of the Finnish universities. During the more recent period (2010-2013), the citation impact decreased and is comparable to the citation impact of universities between ranking position 150 and 175 such as Ghent University or the University of Birmingham. The share of collaborative publications by CoEs (approximately 80%) is lower than for most Finnish universities during 2010-2013, but the share of internationally collaborative publications (60%) and industrial collaborative publications (8%) is higher than for (almost) any Finnish university during this period of analysis. Similar comparisons can be made for the citation impact and scientific collaboration at the level of main fields.

Appendix I. Bibliometric indicators

In this appendix, we describe the methods underlying the present bibliometric analysis.

A1.1. Output indicator

The publication output indicator, denoted by P, measures the total publication output of a research unit. It is calculated by counting the total number of publications of a research unit, including only publications covered by WoS. The counting method is fractional counting unless stated otherwise. We stress that research articles and review articles are the only publication types that are taken into account. Other publication types, such as editorial material, meeting abstracts, letters, and book reviews, are not included.

A1.2. Impact indicators

A number of indicators are available for measuring the scientific impact of the publications of a research unit. These indicators relate to the number of times publications have been cited.

Self-citations

In the calculation of all our impact indicators, we disregard author self-citations. We classify a citation as an author self-citation if the citing publication and the cited publication have at least one author name (i.e., last name and initials) in common. In this way, we ensure that our indicators focus on measuring only the contribution and impact of the work of a researcher on the work of other members of the scientific community. Sometimes self-citations can serve as a mechanism for self-promotion rather than as a mechanism for indicating relevant related work. The impact of the work of a researcher on his own work is therefore ignored.

Counting method

In computing the impact indicators, we apply the fractional counting method, where depending on the co-authorship nature of a publication only a certain fraction of the publication is assigned to the research unit. This is opposed to the full counting method. In case of full counting, publications are always fully assigned to research units, regardless of the collaboration nature of the authorship, e.g., single-authored, two authors from the same research unit, or two or more authors from the same or different countries. Impact indicators calculated using fractional counting tend to have lower values than impact indicators calculated using full counting. This is caused by for the fact that collaborative publications on average tend to have a higher citation impact than non-collaborative publications, which is corrected for in fractional counting. Since fractional counting corrects for differences in collaboration behavior across fields, citation impact results can be compared more fairly across fields. A disadvantage may be, however, that fractional counting is sometimes perceived as less intuitive and more difficult to interpret than full counting.

Non-normalized indicators of citation impact

The total citation score (TCS) indicator gives the total number of citations received by the publications of a research unit. The mean citation score (MCS) indicator equals the average number of citations per publication. This indicator is obtained by dividing TCS by P, the total number of publications.

The PP(uncited) indicator reports the number of uncited publications as a proportion of the total number of publications of a research unit.

Normalized indicators of citation impact

Usually, a recent publication has received fewer citations than a publication that appeared a number of years earlier. Moreover, for the same publication year, publications in for instance mathematics have usually received a much smaller number of citations than publications in for instance biology. This is due to the different citation cultures in different fields. To account for these age and field differences in citations, we use normalized citation indicators. In this case, citation impact indicators have been normalized at the level of the 4,000 subfields (hereafter: fields) from the CWTS publication level classification.

The mean normalized citation score indicator, denoted by MNCS, provides a more sophisticated alternative to the MCS indicator. The MNCS indicator is similar to the MCS indicator except that it performs a normalization that aims to correct for differences in citation characteristics between publications from different fields and between publications of different ages. To calculate the MNCS indicator for a unit, we first calculate the normalized citation score of each publication of the unit. The normalized citation score of a publication equals the ratio of the actual and the expected number of citations of the publication, where the expected number of citations is defined as the average number of citations of all publications (i.e., research articles and review articles) that belong to the same field and that appeared in the same publication year.

The MNCS indicator is obtained by averaging the normalized citation scores of all publications of a unit. If a unit has a value of one for the MNCS indicator, this means that on average the actual number of citations of the publications of the unit equals the expected number of citations. In other words, on average the publications of the unit have been cited equally frequently as publications that are similar in terms of field and publication year. An MNCS indicator of, for instance, two means that on average the publications of a unit have been cited twice as frequently as would be expected based on their field and publication year. We refer to Appendix II for an example of the calculation of the MNCS indicator.

Since the MNCS indicator is based on averages and since citation distributions tend to be highly skewed, the MNCS indicator may sometimes be strongly influenced by a single very highly cited publication. If a unit has one such publication, this is usually sufficient for a high score on the MNCS indicator, even if the other publications of the unit have received only a limited number of citations. Because of this, the MNCS indicator may sometimes seem to significantly overestimate the actual scientific impact of the publications of a research unit.

Therefore, in addition to the MNCS indicator, we use another important impact indicator. This is PP(top 10%), the proportion of the publications of a research unit that belong to the top 10% mostly frequently cited publications in their field and publication year.

For each publication of a research unit, the PP(top 10%) indicator determines, based on the number of citations of the publication, whether the publication belongs to the top 10% of all publications in the same field and the same publication year. The PP(top 10%) indicator equals the proportion of the publications of a research unit that are in the top 10% of their field and publication year. If a research unit has a value of 10% for the PP(top 10%) indicator, this means that the actual number of top 10% publications of the unit equals the expected number. A value of 20% for the PP(top 10%) indicator for instance means that a unit has twice as many top 10% publications as expected. We note that in addition to the PP(top 10%) indicator we also have the P(top 10%) indicator. This indicator equals the number of top 10% publications of a research unit. The P(top 10%) indicator is obtained by multiplying the PP(top 10%) indicator by the P indicator.

To assess the impact of the publications of a research unit, our general recommendation is to rely on the combination of the PP(top 10%) indicator and the MNCS indicator. These two indicators are strongly complementary to each other. The MCS indicator does not correct for field differences and should therefore be used only for comparisons of units that are active in the same field.

Indicators of journal impact

We use the mean normalized journal score indicator (MNJS) to measure the impact of the journals in which a research unit has published. For this, we first calculate the normalized journal score of each publication of the unit. The normalized journal score of a publication equals the ratio of on the one hand the average number of citations of all publications published in the same journal and field and in the same year and on the other hand the average number of citations of all publications published in the same field and the same year. The MNJS indicator is closely related to the MNCS indicator. The difference is that instead of the actual number of citations of a publication, the MNJS indicator uses the average number of citations of all publications published in a particular journal and field. The interpretation of the MNJS indicator is analogous to the interpretation of the MNCS indicator. If a unit has a value of one for the MNJS indicator, this means that on average the unit has published in journals that are cited equally frequent as would be expected based on their field. Likewise, a value of two for the MNJS indicator means that on average a unit has published in journals that are cited twice as frequently as would be expected based on their field.

A1.3. Indicators of scientific collaboration

Indicators of scientific collaboration are based on an analysis of the addresses listed in the publications produced by a research unit. We first identify publications authored by a single institution ('no collaboration'). Subsequently, we identify publications that have been produced by institutions from different countries ('international collaboration') and publications that have been produced by multiple institutions from the same country ('national collaboration'). These types of collaboration are mutually exclusive. Publications involving both national and international collaboration are classified as international collaboration.

Appendix II. Calculation of field-normalized indicators

To illustrate the calculation of the MNCS indicator, we consider a hypothetical research group that has only five publications. Table A1. provides some bibliometric data for these five publications. For each publication, the table shows the scientific field to which the publication belongs, the year in which the publication appeared, and the actual and the expected number of citations of the publication. The expected number of citations is the average number of citations per publication in a particular field. (For the moment, the last column of the table can be ignored.) As can be seen in the table, publications 1 and 2 have the same expected number of citations. This is because these two publications belong to the same field and have the same publication year. Publication 5 also belongs to the same field. However, this publication has a more recent publication year, and it therefore has a smaller expected number of citations. It can further be seen that publications 3 and 4 have the same publication year. The fact that publication 4 has a larger expected number of citations than publication 3 indicates that publication 4 belongs to a field with a higher citation density than the field in which publication 3 was published.

The MNCS indicator equals the average of the ratios of actual and expected citation scores of the five publications. Based on Table A1, we obtain

$$\text{MNCS} = \frac{1}{5}\left(\frac{7}{6.13} + \frac{37}{6.13} + \frac{4}{5.66} + \frac{23}{9.10} + \frac{0}{1.80}\right) = 2.08$$

Hence, on average the publications of our hypothetical research group have been cited more than twice as frequently as would be expected based on their field and publication year.

Table A1. Bibliometric data for the publications of a hypothetical research group.

Publication	Field	Year	Actual Citations	Expected Citations	Top 10% threshold
1	Surgery	2007	7	6.13	15
2	Surgery	2007	37	6.13	15
3	Clinical neurology	2008	4	5.66	13
4	Hematology	2008	23	9.10	21
5	Surgery	2009	0	1.80	5

To illustrate the calculation of the PP(top 10%) indicator, we use the same example as we did for the MNCS indicator. Table A1. shows the bibliometric data for the five publications of the hypothetical research group that we consider. The last column of the table indicates for each publication the minimum number of citations needed to belong to the top 10% of all publications in the same field and the same publication year.[1] Of the five publications, there are two (i.e., publications 2 and 4) whose number of citations is above the top 10% threshold. These two publications are top 10% publications. It follows that the PPtop10% indicator equals

$$PP_{top10\%} = \frac{2}{5} = 0.4 = 40\%$$

In other words, top 10% publications are four times overrepresented in the set of publications of our hypothetical research group.

[1] If the number of citations of a publication is exactly equal to the top 10% threshold, the publication is partly classified as a top 10% publication and partly classified as a non-top-10% publication. This is done in order to ensure that for each combination of a field and a publication year we end up with exactly 10% top 10% publications.