



# CWTS BIBLIOMETRIC REPORT

Meaningful metrics

## Research performance analysis for the SOR programme of the Rijksinstituut voor Volksgezondheid en Milieu (RIVM) 2011-2014/15

*July, 2017*



Universiteit  
Leiden



## **Project team**

**Erik van Wijk**, *Project leader*

**Ingeborg Meijer**

CWTS B.V.

P.O. Box 905

2300 AX Leiden, The Netherlands

Tel. +31 71 527 3948

Fax +31 71 527 3911

E-mail [wijk@cwts.leidenuniv.nl](mailto:wijk@cwts.leidenuniv.nl)

### **General parameters of the bibliometric study**

Database:	All publications in Web of Science Core database
Classification system:	Web of Science journal subject categories
Publication window:	2011-2014
Publication types:	Articles, Review, Letters
Citing publications:	All publication types
Citation window:	Variable length until and including 2015
Letters:	Included (weight 0.25)
Counting method:	Wherever possible whole counting
Self-citations:	Excluded
Top indicators:	Top 10%

### **Acknowledgements**

CWTS wishes to thank RIVM for supplying CWTS with machine readable input for the data collection.

# Table of contents

<b>1. Introduction.....</b>	<b>7</b>
1.1. Goal of the research .....	7
<b>2. Data collection and bibliometric indicators .....</b>	<b>9</b>
2.1. Bibliometric indicators overview .....	9
<b>3. Results .....</b>	<b>11</b>
3.1. SOR Overall research profile analysis .....	11
3.2. SOR General research profile analysis by scientific field.....	13
3.3. SOR Spearhead level analysis and research profile .....	16
3.4. SOR programme collaboration .....	28
<b>4. Main findings.....</b>	<b>30</b>
<b>Appendix I. Data collection, selection and handling .....</b>	<b>32</b>
Initial database structure .....	32
Bibliometric approach.....	32
Coverage of publications .....	32
<b>Appendix II. Bibliometric indicators.....</b>	<b>35</b>
II.1. General matters.....	35
II.2. Output indicator .....	35
II.3. Impact indicators .....	36
Self-citations .....	36
Counting method.....	36
Un-normalized indicators of citation impact.....	36
Normalized indicators of citation impact.....	36
Publications belonging to multiple fields .....	38
Limitations of field normalization.....	38
Indicators of journal impact .....	39

II.4. Indicators of scientific co-operation .....	39
<b>Appendix III. Calculation of field-normalized indicators .....</b>	<b>40</b>
<b>Appendix IV. Overview underlying data files.....</b>	<b>42</b>

# 1. Introduction

The Rijksinstituut voor Volksgezondheid en Milieu (RIVM) has a specific budget for its Strategic Research Programme (SOR). This research is aimed at providing RIVM with the required (scientific) expertise and quality, so that it can perform its tasks effectively both now and in the future, thus contributing to a healthy population and a healthy human environment. It hence is intended to help anticipating to future research questions. Also, the SOR budget allows participating in international research networks. SOR projects are carried out in a four-year cycle.

The report 'Evaluation RIVM Strategic Research 2011-2014 - Ready for the challenges of tomorrow' (RIVM2015-0103) describes the results of an evaluation of RIVM's Strategic Research Programme (SOR) for the 2011-2014 period. During the period surveyed, a total of 107 projects were carried out at a total cost of approximately 45 M€. 17% of the total number of RIVM publications in the period 2011-2014 is connected to SOR-projects, whereas on average 5% of the RIVM-budget is dedicated to SOR projects. This suggests that the Strategic Research Programme significantly contributes to the total number of publications of RIVM.

The Strategic research Programme has selected seven spearheads ('speerpunten') to structure the programme. The spearheads were selected in 2009, with a view on the most important strategic questions in health and environment issues. They were chosen in deliberation with external experts that assign projects to RIVM and the Supervisory Committee of RIVM. Projects within the spearheads were selected bottom up, and the evaluation was based on quality, creativity, and potential for use. Many RIVM research is multidisciplinary in nature, and this is particularly true for spearhead projects. Spearhead coordinators have to maintain the focus within the spearheads.

The reason to conduct the current study is that RIVM applied its own methodology for measuring the scientific quality of SOR that is not fully compliant with CWTS' most recent insights in bibliometric analyses. Therefore, the analysis has been carried out according to the current bibliometric performance assessment standards. These are described throughout the report and in the annex when information is too detailed.

## 1.1. Goal of the research

The Rijksinstituut voor Volksgezondheid en Milieu (RIVM) has requested the Centre for Science and Technology Studies (CWTS) of Leiden University to perform this bibliometric analysis. The goal of the project is to gain concrete and detailed insight into the bibliometric performance of

the research publications of SOR's Strategic Research Programme. To this end CWTS performed three types of analyses.

- 1) Overall research profile analysis (3.1).
- 2) General research profile analysis (3.2)
- 3) Research profile analysis at Spearhead level (3.3)
- 4) Collaboration analysis (3.4)

The results of the analysis performed by CWTS are presented in this report.

Our report focuses on the publication output of SOR during 2011-2014. The citation impact of these publications is measured during the time period with one year added to allow 2014 publications to also gather citations and is compared to worldwide reference values. The study is based on a quantitative analysis of scientific articles, reviews and letters; published in international journals covered by the Web of Science (WoS), a publication database used and enhanced by CWTS.

The objective of the analysis is to assess the publication activity and international impact of SOR publications, the publication profiles of the Institute's SOR and its spearheads within different areas of research and the collaboration in the national and international context.

Before presenting the actual findings of the analyses, in the next paragraph we shortly introduce a definition of bibliometric terms used within the report.



## 2. Data collection and bibliometric indicators

Data acquisition is a crucial step in any bibliometric analysis. It determines to some extent the value and the meaning of the statistics that are calculated. The full details of the procedure and methodology are described in Appendix I. The results indicate a good overall coverage (86%) for SOR publications as a whole and on the level of the Speerpunten. The internal coverage provides insight into the citing practices of SOR. When internal coverage percentage drops below 50% it is not possible to perform robust analyses with confidence as this is an indication that the non-WoS citation environment is as important as the environment within WoS. In this case the results indicate a good overall coverage for SOR publications as a whole and on the level of the Speerpunten. 14% of the documents cited by the SOR articles, reviews and letters are published in sources not covered by WoS. This could reflect citations to e.g. books and book chapters, conference papers, reports and other grey literature not listed in the WoS.

### 2.1. Bibliometric indicators overview

In order to allowing to compare different scientific fields with different citation behaviour to each other the impact indicators are normalized. E.g. in mathematics there is lower citation practice than in life sciences; this is corrected by normalization. The normalization method by which the impact indicators are normalized is extensively described in Appendix II and II. In short, the key elements for normalization is done on the basis of allocation of publications to scientific field clusters using their inter citation relations. Furthermore in the scientific field analysis we used the WoS category labels but the normalization was still done as stipulated above. Also we reallocated papers within the “Multidisciplinary” category label to better represent the actual fields active in. It is important to emphasize that the correction for field differences that is performed by the MNCS and PPtop10% indicators is only a partial correction. These indicators are based on the field definitions provided by the WoS subject categories. It is clear that fields in reality do not have well-defined boundaries.

Table 1 provides an overview of the CWTS bibliometric indicators. The indicators below are grouped by dimension. More in depth information is provided in Appendix II and Appendix III describing the bibliometric practices in depth.

Table 1.1 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.
PPtop10%	Impact	Proportion of publications that belong to the top 10% of their field. The 'visibility'-index as highly cited work tends to be noted more.
PPnC	Impact	Proportion of uncited publications.
MNJS	Journal impact	Average normalized citation impact of a journal.
PP (Collab)	Collaboration	Proportion of publications resulting from collaboration.
PP (Int Collab)	Collaboration	Proportion of publications resulted from international collaboration.
PP (Industry)	Collaboration	Proportion of publications resulted from a collaboration with an industry partner.

In this report, the following indicators will be provided for each unit of analysis: P, TCS, MCS, TNCS, MNCS, PPtop10%, PPnC, and MNJS.

Indicators based on a limited volume of publications are to be viewed with some caution. The robustness is indicated by the corresponding levels of the MNCS (which is influenced by outliers) and the PPTop10%. A very high MNCS in combination with a very low PPTop10% is not a robust situation, but when both are high the results can be considered robust.

## 3. Results

In this section, the results of the performance analysis are reported. Section 3.1 shows the overall results on the level of the SOR and its Speerpunten, 3.2 focuses on scientific fields and 3.3 analyses collaboration types.

Using bibliometric techniques, the present study analyses the publication output from 2011 to 2014 and citation impact of these publications up to 2015. The impact, as measured by normalized citation impact (author self-citations excluded), should be understood within the context of the worldwide reference value which is always 1 or 10% in the case of the PPTop10%. CWTS uses the MNCS impact indicator as a (bibliometric) proxy for quality. However, there will never be a clear cut proof of the relation between bibliometric impact and quality of the work since quality is in itself an arbitrary notion. The results of output and impact analyses from different angles for the SOR are presented in this chapter.

### 3.1. SOR Overall research profile analysis

The performance indicators of total RIVM's SOR publications as well as when they are analysed under the denominator of the Spearheads provide an overview of the performance within the given period of time. Table 3.1 shows the relevant values. Below, CWTS interprets and highlights relevant findings of the overall SOR analysis (top row of table 3.1). The individual spearheads are discussed in more detail in 3.3.

Year	Analysis Unit	P	MCS	TCS	MNCS	MNJS	TNCS	PP (top 10%)	PP (uncited)	Proportion self citations
2011 - 2014	SOR Total	195.75	9.46	1852.25	1.97	1.59	386.08	18%	15%	24%
	Speerpunten									
2011 - 2014	ANT	21.00	17.81	374.00	2.51	1.82	52.78	31%	5%	25%
2011 - 2014	FKA	12.00	3.00	36.00	0.76	1.11	9.13	8%	17%	23%
2011 - 2014	HEA	42.00	5.67	238.00	1.34	1.53	56.26	19%	17%	21%
2011 - 2014	HSL	20.25	6.37	129.00	1.20	1.32	24.38	5%	11%	25%
2011 - 2014	IDD	47.50	13.61	646.25	2.55	1.83	120.91	22%	24%	23%
2011 - 2014	IRA	26.00	10.23	266.00	4.05	2.10	105.38	27%	4%	21%
2011 - 2014	SVR	28.00	6.25	175.00	0.83	1.03	23.27	7%	14%	26%

Table 3.1 Performance indicators for SOR 2011-2014/15.

From this table can be observed that the total of 196 SOR publications (P) have received, on average, 9.46 citations (MCS) in the period 2011-2014/15. The citation impact is almost twice World Average, with an MNCS of 1.97 (1 is average). Furthermore, 18% of SOR publications belong to the top 10% highly cited publications. CWTS uses this indicator as a measure of 'visibility', since very highly cited papers tend to generate attention within the scientific community. For SOR this means that they are very highly noticed and their work is frequently used in further scientific research. With respect to the journals in which SOR publications appear, it can be concluded that these journals also have an impact value exceeding World Average, since MNJS is 1.59. Finally, around 15% of SOR publications are uncited. Not visible in this table but shown in the accompanying tables underlying this report is that the level of uncitedness is around 6% except for the last publication year 2014 for which there is limited citation information available and when it shoots up to 26%. Of course in later years limited citations are available.

SOR publications perform roughly at 2 times World Average on two of the most important indicators: the MNCS and the PP Top 10%. . The MNCS shows a value that corresponds to a high PPTop 10% indicating that the outcomes are robust. The publications are published in highly cited journals that outperform average world level by a factor 0,6. The level of publications not cited at all is low. SOR produces highly visible, (inter)nationally acclaimed scientific research in internationally highly acknowledged journals. The impact level is not largely carried by very few SOR publications but there is a stable citation contribution across the oeuvre.

There are noticeable differences between the different research publications when we group them according to Speerpunten. 'IRA', 'ANT', 'IDD' score a high to very high MNCS. Whereas 'SVR' and 'FKA' do not perform on World Average level. As stated previously we must bear in mind the limited number of publications the analyses for the Speerpunten are based upon.

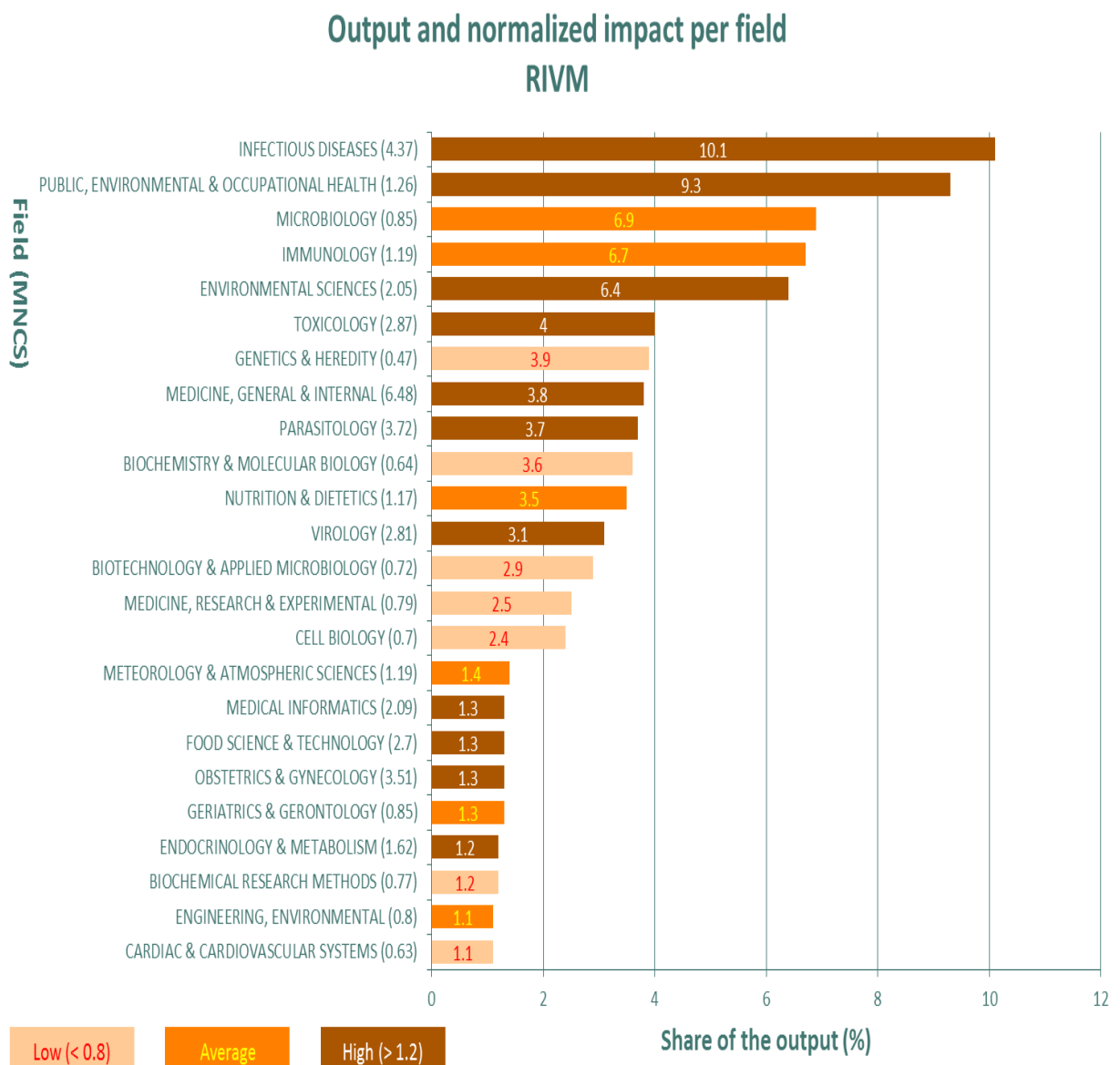
For 'FKA' for example with 12 publications the addition of one highly cited paper would probably influence the total statistics significantly. Not however the PP Top 10%. This indicator is robust by nature and is also below World Average. 'HSL' is an interesting data-point as well. Here the MNCS is high (20% above World Average) but the PP Top 10% is only 5%. This inconsistency shows that although publications of this Speerpunt are cited well overall, there is little top cited work and visibility is therefore not that high.

All in all, SOR scores are high, be it that the differences between the Speerpunten is high as well, ranging from 15% under World Average to over 4 times World Average. This indicates that the overall result is not carried equally by the different research within the SOR themes.

### 3.2. SOR General research profile analysis by scientific field.

The SOR performs research within the institute that is geared towards environment and health related issues. The health and environment research is carried out in different scientific fields of research. These research focal points hence show up as a wide variety of Web of Science research categories. It shows the breadth in which the SOR research is active. CWTS analysed the performance within these different scientific fields. We present the scientific fields and the MNCS impact for those fields with a share of 1% or more in the total output in figure 3.1.

Figure 3.1 Impact most important scientific fields SOR 2011-2014/15



Strikingly, there is no scientific field within the top selection with a share of 4% or more (production P) that have a 'low' label for scientific impact. They all qualify predominantly as 'high' (MNCS above 120%). No upper boundaries are defined for the MNCS impact. However, an MNCS above 2 can be qualified as 'very high'. SOR research has 'high' citation impact for almost half of the scientific fields with a share of 1% or more in the total output.

The scientific fields that comprise the top of the distributions with a share of more than 1% are predominantly environmental and medical health related issues. "Infectious diseases" taking the lead closely followed by "Public, Environmental & Occupational Health", both with very high impact. But the extremely high impact of "Infectious Diseases" should be understood against the backdrop of a total number of publications within that category of not yet 20. There is also an indication that this extremely high impact figure is not a coincidence because of the PP Top 10% indicator at 28% (not shown in the figure but available in the underlying data accompanying this report) . This indicates that the impact of the MNCS is structural on the basis of a coinciding indicator that's much less susceptible for outliers.

Distributing these fields in three main groups is shown in figure 3.2. In the upper table of this figure, all fields with a high MNCS (>1,20) together represent 50% of all SOR publication output. These fields can be considered the core of RIVM research. The second table shows that less than 20% of all output is below world average. This category includes microbiology and applied microbiology, as well as biochemistry and molecular biology. As for the fields with the high MNCS this could be caused by a low number of publications. The third table is a summary of patches of research with a high MNCS but small output (together around 5%). As for all other fields it could be caused by an outlier, but it may also represent new clusters of research that relate to specific spearheads that are of good quality.

In the next paragraph the specific spearheads are described in more detail.

Figure 3.2 Main groups of research

Field	MNCS high	Share of output
Medicine, general & internal	6,48	3,8
Infectious diseases	4,37	10,1
Parasitology	3,72	3,7
Toxicology	2,87	4
Virology	2,81	3,1
Environmental sciences	2,05	6,4
Public, environmental & occupational health	1,26	9,3
Immunology	1,19	6,7
Nutrition & dietetics	1,17	3,1
<b>Total</b>		<b>50,2</b>

Field	MNCS low	Share of output
Microbiology	0,85	6,9
Biotechnology & applied microbiology	0,72	2,9
Biochemistry & molecular biology	0,64	3,6
Genetics & heredity	0,47	3,9
<b>Total</b>		<b>17,3</b>

Field	MNCS high	Share of output
Obstetrics and gynecology	3,52	1,3
Food science & technology	2,7	1,3
Medical informatics	2,09	1,3
Endocrinology & metabolism	1,62	1,2
<b>Total</b>		<b>5,1</b>

### 3.3. SOR Spearhead level analysis and research profile

1. The Speerpunt **Application of New Technologies (ANT)** consisted of 11 projects. Key words for ANT are: innovation, home & personal care, e-health and IT, sensor diagnostics, personalized medicine, biomaterials, nanotechnology. Compared to the SOR total, the ANT speerpunt has an even higher MNCS than SOR total and a high PPTop10% as well, indicating robustly that this spearhead has performed 2,5 times higher than world average and is very visible as well.

Year	Analysis Unit	P	MCS	TCS	MNCS	MNJS	TNCS	PP (top 10%)	PP (uncited)	Proportion self citations
2011 - 2014	RIVM Total	195.75	9.46	1852.25	1.97	1.59	386.08	18%	15%	24%
	Speerpunten									
2011 - 2014	ANT	21.00	17.81	374.00	2.51	1.82	52.78	31%	5%	25%

In the textbox a selection of funded projects in ANT is presented. It shows a wide variety of topics, from which it is not directly clear to which research fields the resulting publications 'belong'. Analysis of the ANT research fields in figure 3.3 shows that it is mainly 'infectious diseases', 'parasitology', and 'obstetrics & gynaecology' that score high. This could be related to

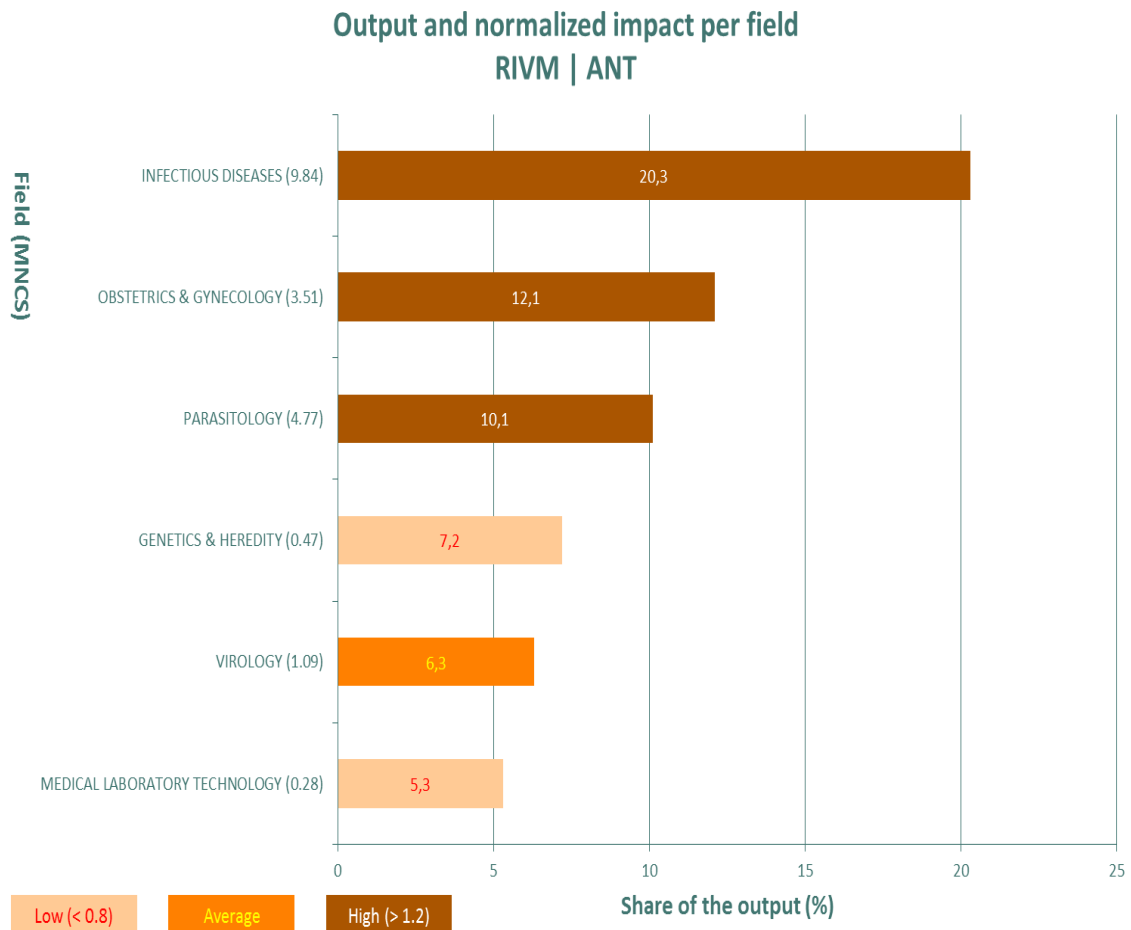
#### *Projects in Application of new technology*

- *Het gebruik van proteomics voor bevolkingsonderzoeken en diagnostiek*
- *Het gebruik van menselijke stamcellen als alternatief voor dierproeven*
- *De impact van medische technologie op gezondheid en zorguitgaven*
- *Een nieuwe methode voor het meten van micro-organismen die door de lucht worden verspreid*
- *Innovatieve methoden om luchtverontreiniging te meten*
- *Een nieuwe statistische methode om uitbraken van infectieziekten te meten*



the “Bevolkingsonderzoek” research, but the common denominator of the ANT spearhead does not show from the fields. Notably, there are some low-scoring fields in this spearhead as well.

Figure 3.3. ANT spearhead output and MNCS



2. The Speerpunt **Filling the gap: from Knowledge to Action (FKA)** consisted of 12 projects. Key words for FKA are risk perception, communication, interactive websites, behavioural change, implementation, knowledge management, stakeholder engagement, societal impact. Compared to the SOR total, the FKA speerpunt has a considerable lower MNCS than SOR total, indicating that this spearhead has performed lower than world average. Also the PPTop10% is not that high (8%) indicating that the visibility is lower than average as well. Although the number of publications in this speerpunt is low, the internal coverage is still 75% (see appendix I).

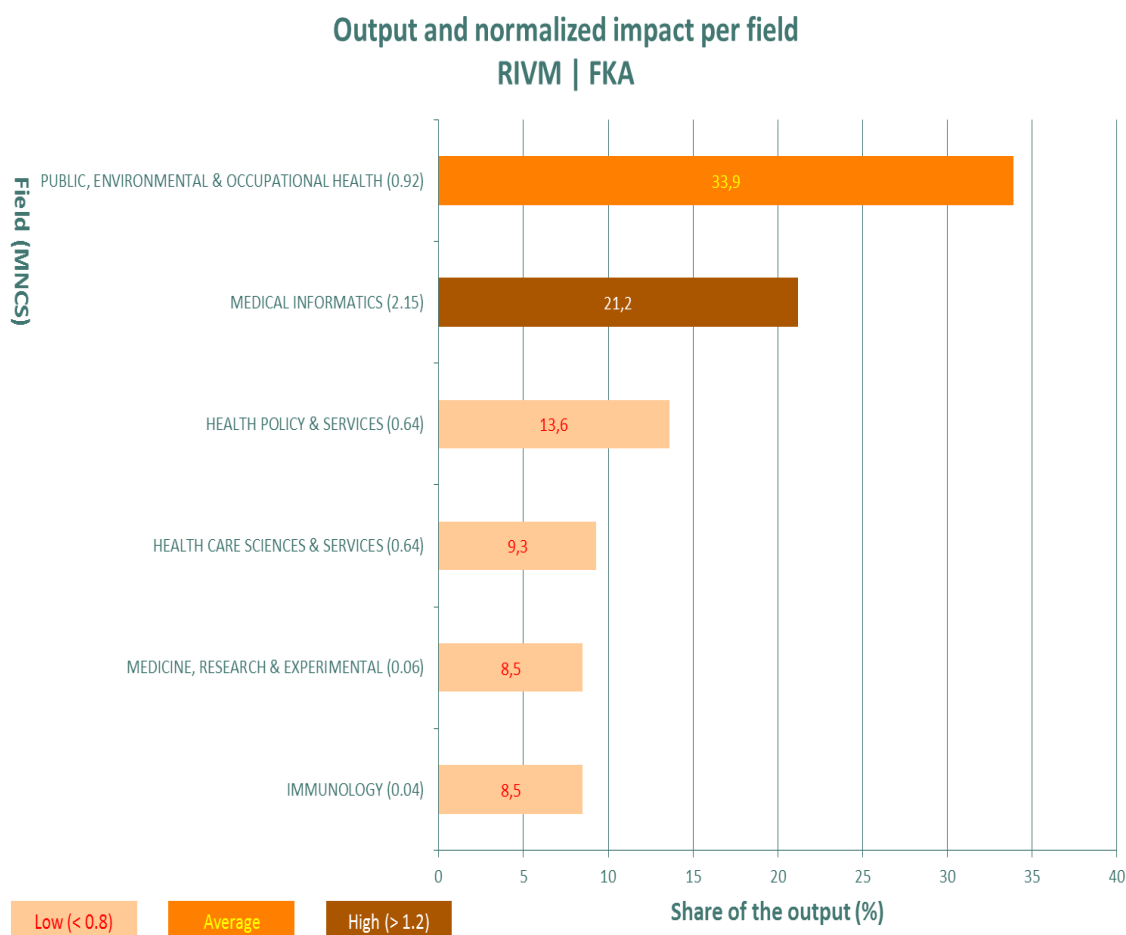
Year	Analysis Unit	P	MCS	TCS	MNCS	MNJS	TNCS	PP (top 10%)	PP (uncited)	Proportion self citations
2011 - 2014	RIVM Total	195.75	9.46	1852.25	1.97	1.59	386.08	18%	15%	24%
	Speerpunten									
2011 - 2014	FKA	12.00	3.00	36.00	0.76	1.11	9.13	8%	17%	23%

In the textbox a selection of funded projects in FKA is presented. It shows a variety of topics, from which 'communication' is a common theme. It is however not directly clear to which research fields the resulting publications 'belong'. Analysis of the FKA research fields in figure 3.4 shows that it is mainly 'public, environmental & occupational health', and 'medical informatics'. Both align with keywords and projects in this speerpunt.

*Projects in Filling the gap: from knowledge to action*

- *Communicatie over vaccinatie bij pandemieën*
- *Verbeteren van de communicatie voor het verspreiden van kennis over preventie van ziekten*
- *De betekenis van 'health literacy' voor de volksgezondheid*
- *Het opzetten van een systeem om de redenen van ouders om hun kinderen wel of niet te laten vaccineren en de houding van de consultatiebureau-medewerkers te monitoren*

Figure 3.4 FKA spearhead output and MNCS



Notably, there are some low-scoring fields in this spearhead as well, relating to health policy and health sciences mainly. This theme is less technological driven than e.g. ANT and could be more focused toward other, more societal products than on scientific publications. However, in the RIVM evaluation report of the SOR 2011-2014 (RIVM report 2015-0103 table 9 and table 10), the FKA theme doesn't produce more 'other products (e.g. presentations or reports)' or societal impact (e.g. use in guidelines or international committees).

3. The Speerpunt **Healthy Ageing (HEA)** consisted of 19 projects and aligns with the European grand societal challenges. Key words for HEA are life style and risk factors, healthy food & nutrition, occupational and environmental health, epidemiology, antibiotic resistance, hospital infections, fragile elderly, chronic diseases, alcohol and drugs, multi morbidity. Compared to the SOR total, the HEA speerpunt with its 42 publications has a lower MNCS than SOR total but is still high and above world average (1,34). The PPTop10% however is equally high as SOR total showing a robust visibility of this topic.

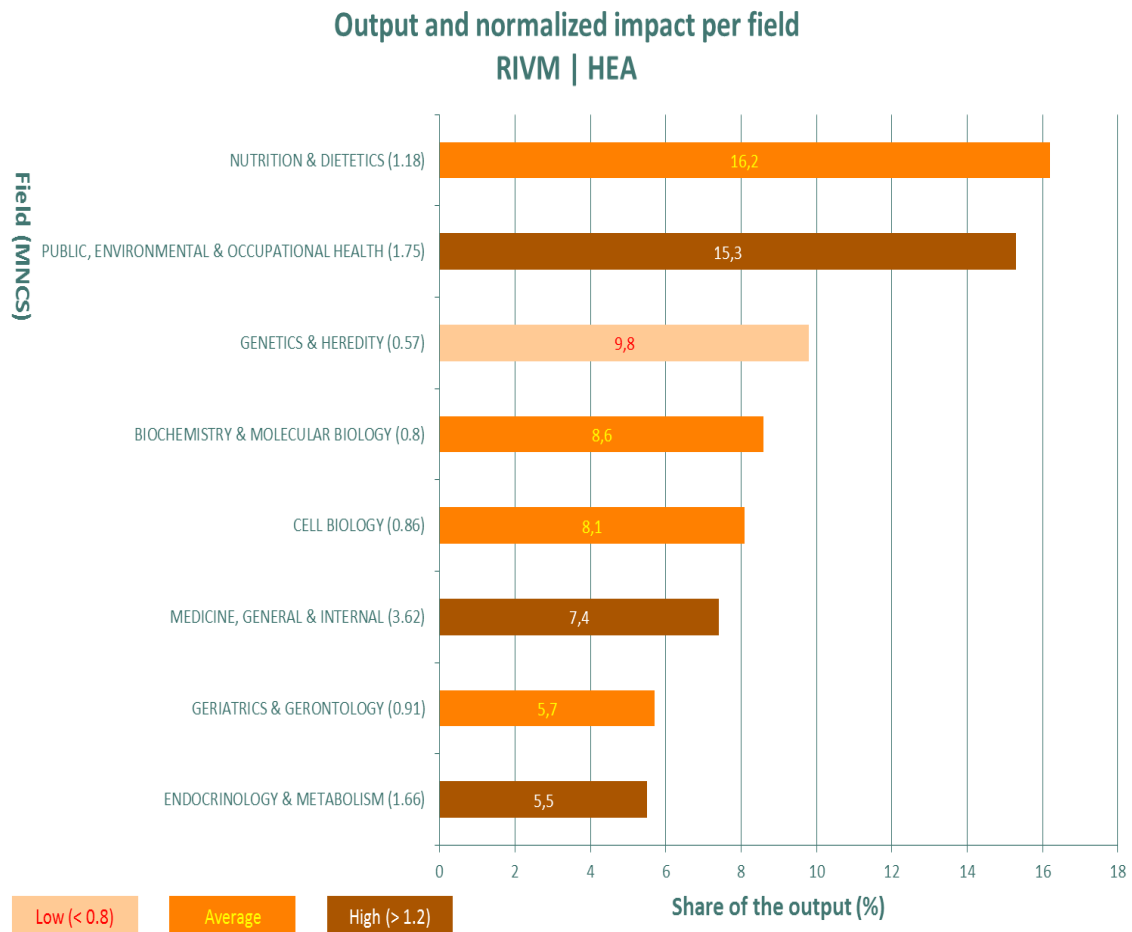
Year	Analysis Unit	P	MCS	TCS	MNCS	MNJS	TNCS	PP (top 10%)	PP (uncited)	Proportion self citations
2011 - 2014	RIVM Total	195.75	9.46	1852.25	1.97	1.59	386.08	18%	15%	24%
Speerpunten										
2011 - 2014	HEA	42.00	5.67	238.00	1.34	1.53	56.26	19%	17%	21%

In the textbox a selection of funded projects in HEA is presented. It shows a variety of topics, ranging from societal participation of elderly to cellular aspects of ageing. Likewise, there is a broad range of research fields that 'belong' to this spearhead (see figure 3.5). The most productive ones, both with an MNCS above world average are 'public, environmental & occupational health', and 'nutrition & dietetics'. Both align with keywords and projects in this spearpunt. The other research fields in this spearhead are below world average with the exception of 'medicine, general & internal'.

*Projects in Healthy ageing*

- *Factoren die van invloed zijn op maatschappelijke participatie van ouderen*
- *Mutaties in het DNA en veroudering van cellen*
- *Effecten van invloeden tijdens de zwangerschap op latere ziekten*
- *Een internationaal overzicht van chronische ziekten*

Figure 3.5 HEA output and MNCS



4. The Speerpunt **Healthy and Sustainable Living environments (HSL)** consisted of 11 projects and aligns with the European grand societal challenges. Keywords are local living environment, ecosystem, energy, animal welfare, consumers, climate change, CO<sub>2</sub>- balance, long-term effects, cost/benefit. Compared to the SOR total, the HSL speerpunt with its 20 publications has a lower MNCS than SOR total but is still high and above world average (1,20). The PPTop10% however is below 10%, indicating that this research topic is not so visible.

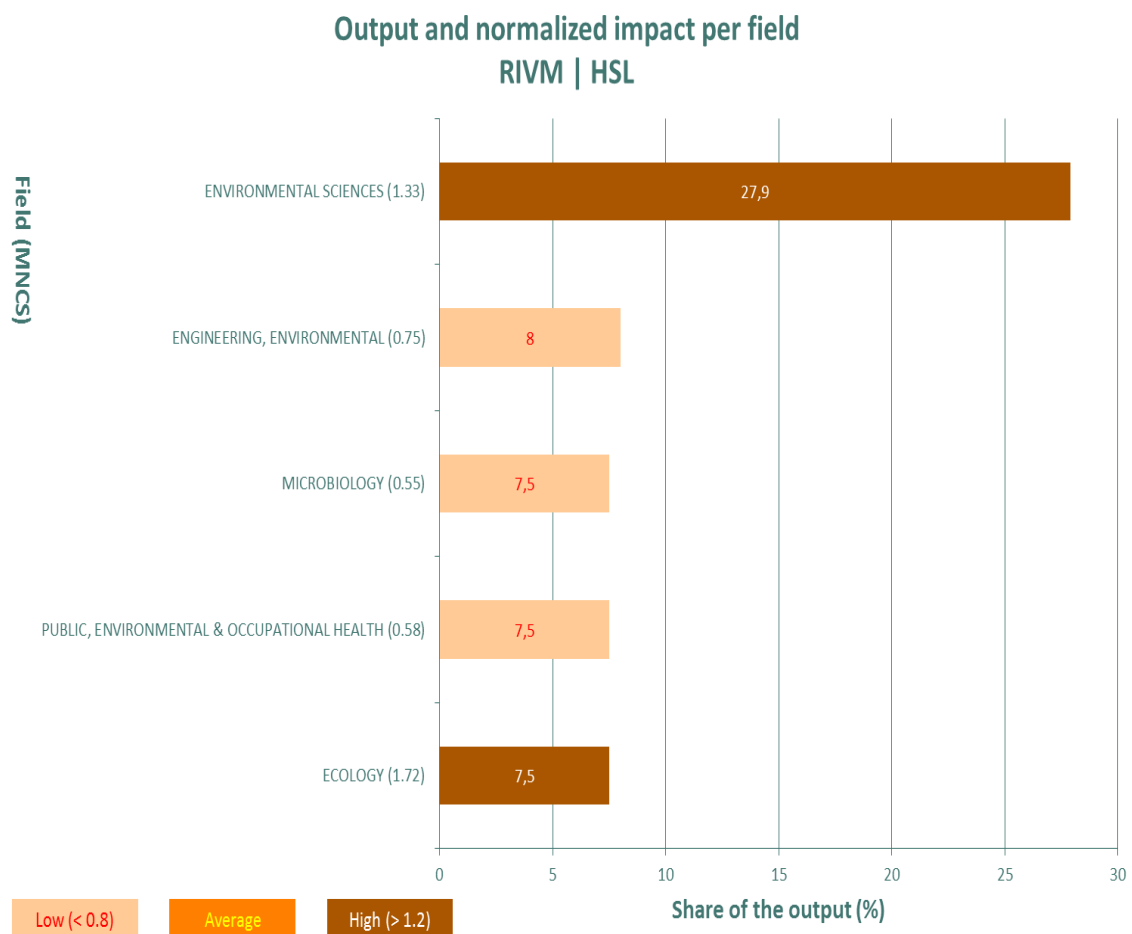
Year	Analysis Unit	P	MCS	TCS	MNCS	MNJS	TNCS	PP (top 10%)	PP (uncited)	Proportion self citations
2011 - 2014	RIVM Total	195.75	9.46	1852.25	1.97	1.59	386.08	18%	15%	24%
2011 - 2014	Speerpunten									
2011 - 2014	HSL	20.25	6.37	129.00	1.20	1.32	24.38	5%	11%	25%

In the textbox, a selection of funded projects in HSL is presented. It shows predominantly environmental topics. Likewise, the main research fields are ‘environmental sciences’ and ‘ecology’, both with an MNCS a little above world average (see figure 3.6). The other research fields in this spearhead are below world average. No research fields score average.

*Projects in Healthy and sustainable living environments*

- *Een nieuwe meetmethode om straling van zendapparatuur te meten*
- *Humane enterovirussen en parechovirussen in Nederlands riool- en oppervlaktewater: levert dit een gevaar op voor de volksgezondheid?*
- *Lichtvervuiling en de afwezigheid van donkerte*
- *PHENOTYPE: Onderzoek naar de invloed van natuur op gezondheid*
- *Op weg naar een duurzame akoestische leefomgeving*

Figure 3.6 HSL output and MNCS



5. The Spearpunt **Infectious Disease Dynamics (IDD)** consisted of 18 projects. This spearhead reflects the long term focus of RIVM activities and presumably builds upon a solid knowledge base. Keywords are pathogen, host, zoonose, vaccination, resistance, immunology, food related infections, prescription behaviour of antibiotics, prevention strategy. Compared to the SOR total, the IDD spearpunt has an even higher MNCS than SOR total and a high PPtop10% as well, indicating that this spearhead has performed 2,5 times higher than world average and is very visible as well.

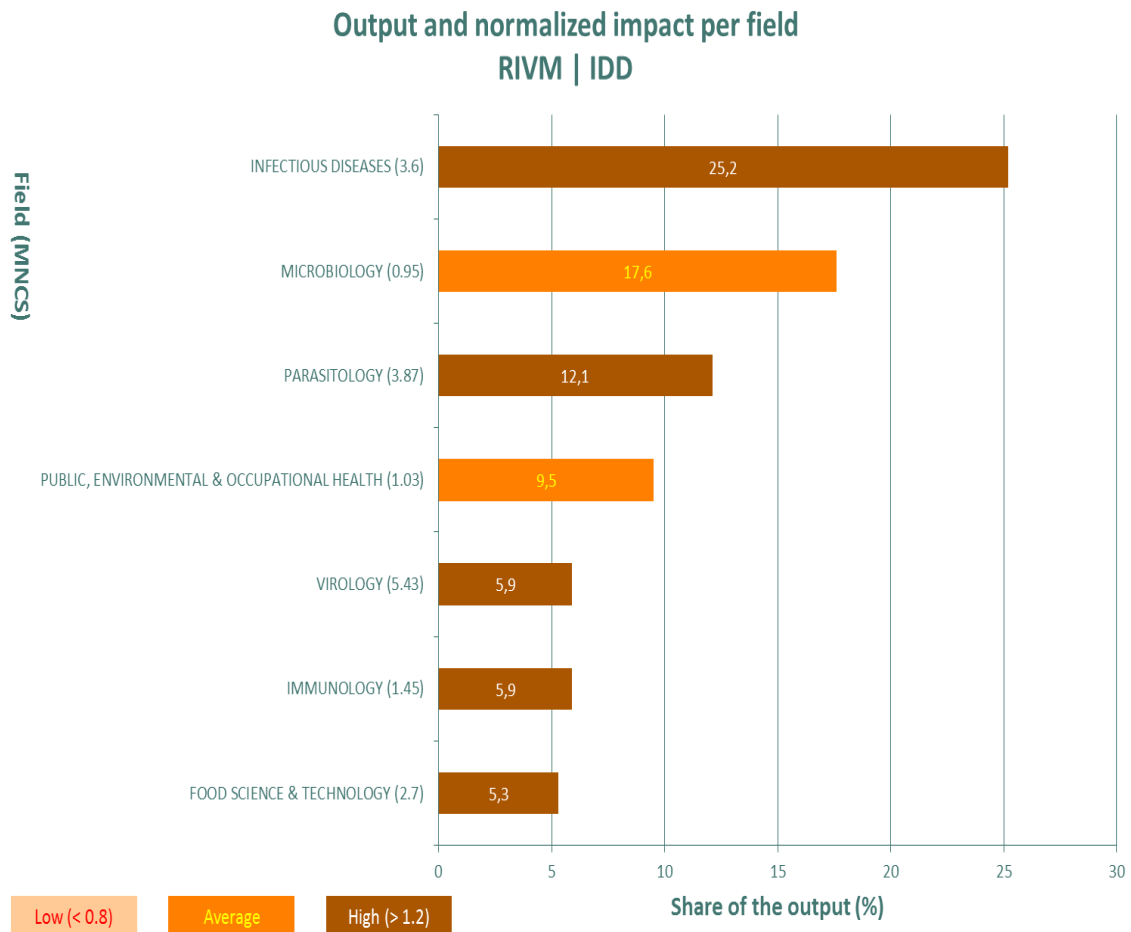
Year	Analysis Unit	P	MCS	TCS	MNCS	MNJS	TNCS	PP (top 10%)	PP (uncited)	Proportion self citations
2011 - 2014	RIVM Total	195.75	9.46	1852.25	1.97	1.59	386.08	18%	15%	24%
	Speerpunten									
2011 - 2014	IDD	47.50	13.61	646.25	2.55	1.83	120.91	22%	24%	23%

In the textbox a selection of funded projects in IDD is presented. It shows a wide variety of topics, predominantly focusing on spreading of infectious diseases. Analysis of the IDD research fields in figure 3.7 shows that it is mainly 'infectious diseases', 'parasitology', and 'virology' that score high. This reflects the core research in the spearhead. Notably, none of the research fields in IDD score below world average. However, the microbiology research field that for SOR overall is relatively low scoring, accounts for nearly one-fifth of the output in this high-scoring spearhead.

*Projects in Infectious diseases dynamics*

- *Biomarkers voor het verloop van Q-koorts*
- *De route van Salmonella besmetting in varkensvlees*
- *Bestrijding van polio met antivirale middelen*
- *Antibioticumresistente bacteriën op groenten*
- *Screening van migranten op hepatitis B en C*
- *Verspreiding van influenza A in de Nederlandse bevolking*
- *Europese samenwerking ter bestrijding van antibioticaresistentie*

Figure 3.7 IDD output and MNCS



6. The Speerpunt **New dimensions on Integrated Risk Assessment in public health and environment (IRA)** consisted of 27 projects, the highest number in a spearhead. This spearhead also reflects a long term focus of RIVM activities and presumably builds upon a solid knowledge base. Key words are modeling, food and nutrition, microbiology, health foresight (VTV), pharmaceuticals, health technology assessment (HTA), new threats, new therapies, quantitative risk assessment, instruments for environmental impact reports. Compared to the SOR total, the IRA spearpunt has a two times higher MNCS than SOR total (which already is twice the world average) and a very high PPTop10% as well (more than one-quarter of the publications is highly visible). Based upon 26 publications (lower than e.g. HEA and IDD) this spearhead has the highest SOR spearpunt performance. This could however be due to a few highly cited publications.



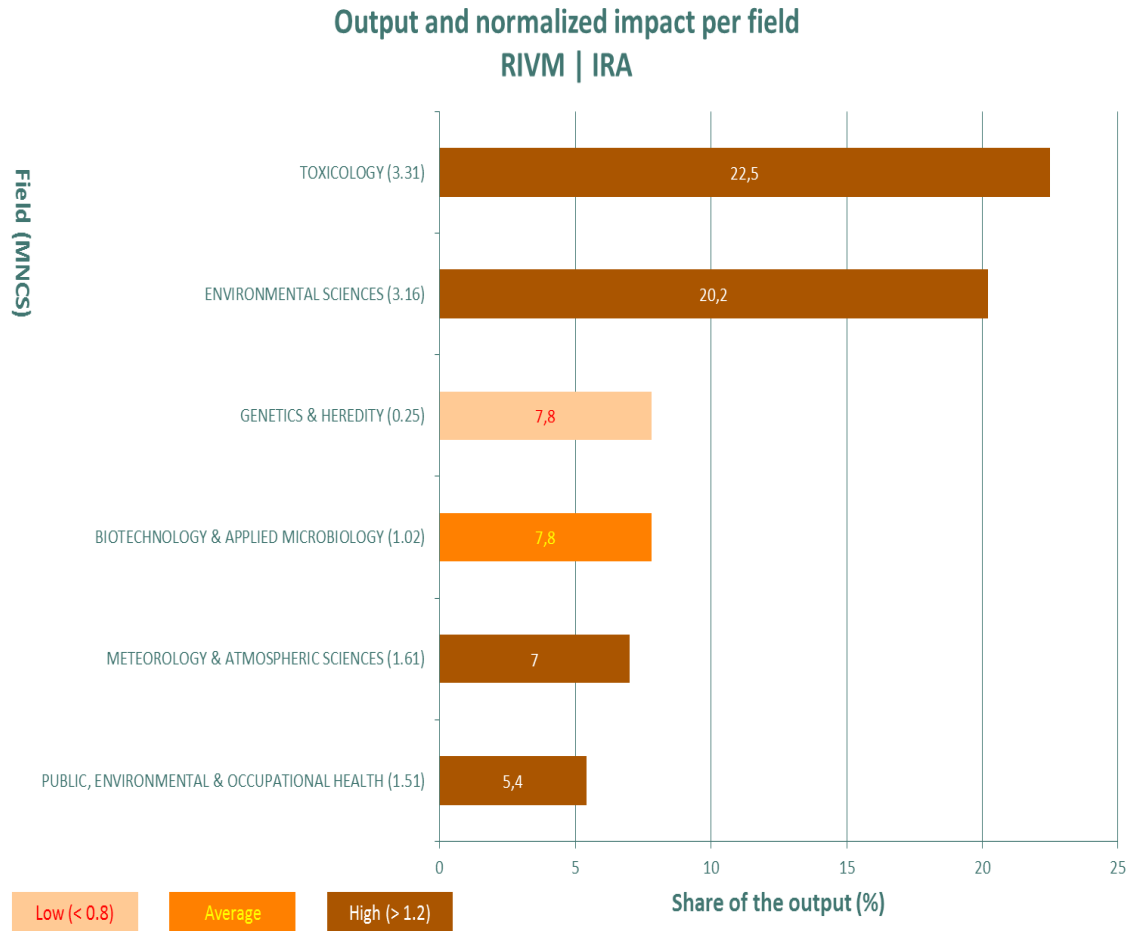
Year	Analysis Unit	P	MCS	TCS	MNCS	MNJS	TNCS	PP (top 10%)	PP (uncited)	Proportion self citations
2011 - 2014	RIVM Total	195.75	9.46	1852.25	1.97	1.59	386.08	18%	15%	24%
	Speerpunten									
2011 - 2014	IRA	26.00	10.23	266.00	4.05	2.10	105.38	27%	4%	21%

In the textbox a selection of funded projects in IRA is presented. It shows a variety of topics, predominantly focusing on (environmental) risk assessment and modelling. Analysis of the IRA research fields in figure 3.8 shows that it is mainly 'toxicology', 'environmental sciences', and 'public, environmental & occupational health' that score a high MNCS. This reflects the core research in the spearhead. Only the 'Genetics & heredity' research field in IRA scores well below world average (MNCS 0,25).

*Projects in New dimensions on Integrated Risk Assessment in public health and environment*

- *Karakterisering van hypergevoeligheid voor milieufactoren*
- *Nieuwe methoden voor risicobeoordeling met behulp van 'omics'*
- *Complexe gezondheidsproblemen ontrafelen via systeemdenken*
- *Een risicobeoordelingsmethode voor toxiciteit van nanomaterialen*
- *HEALTHY ACTION Gezonder leven in een gezonde omgeving?*
- *Effecten van straling op hart en vaten*
- *Modellen om het gedrag van stoffen in het lichaam te voorspellen; nuttig bij het beoordelen van calamiteiten*
- *Stappen naar persoonsgebonden borstkankerscreening*

Figure 3.8 IRA output and MNCS



7. The Speerpunt **Strategic Vaccine Research (SVR)** consisted of 9 projects. In the period under evaluation, the vaccine research of the former Dutch Vaccin Institute (NVI) merged with the Strategic research programme. Later on, the projects related to vaccine technology and innovative vaccine concepts were relocated to the newly starting (2013) Institute for Translational Vaccinology (Intravacc). Main focus of this spearhead is therefore vaccine immunology. The 9 projects resulted in 28 publications with an overall of MNCS at the lower boundary of world average (but close to below world average). Also the PPTop10% is below world average indicating that this research is not so visible.

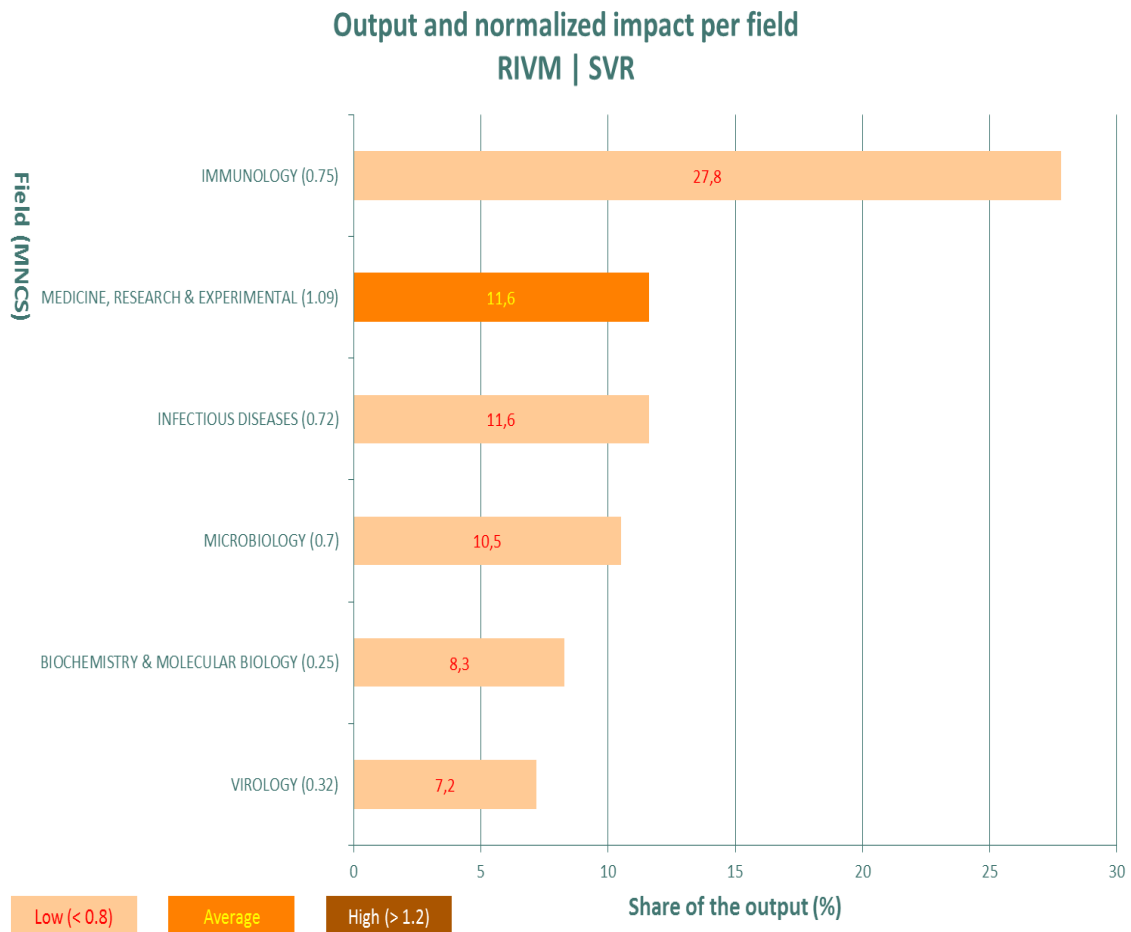
Year	Analysis Unit	P	MCS	TCS	MNCS	MNJS	TNCS	PP (top 10%)	PP (uncited)	Proportion self citations
2011 - 2014	RIVM Total	195.75	9.46	1852.25	1.97	1.59	386.08	18%	15%	24%
	Speerpunten									
2011 - 2014	SVR	28.00	6.25	175.00	0.83	1.03	23.27	7%	14%	26%

In the textbox a selection of funded projects in SVR is presented, focusing on vaccine related research. Analysis of the SVR research fields in figure 3.9 shows that it is – not surprisingly – mainly ‘Immunology’ with 30% of the output. All research fields in this spearhead have an MNCS that is below world average with the exception of ‘medicine, research & experimental’, indicating that the SVR publications attract less citations than SOR total. The lower scores in this spearhead may have been caused by the internal reorganisations.

*Projects in Strategic Vaccine research*

- *Immunological programming*
- *Innovative synthetic vaccines*
- *Identifying long term specific pathogen immunity after vaccination*
- *Innate immunity receptors*
- *T-cells in mumps vaccine failure*

Figure 3.9 SVR output and MNCS



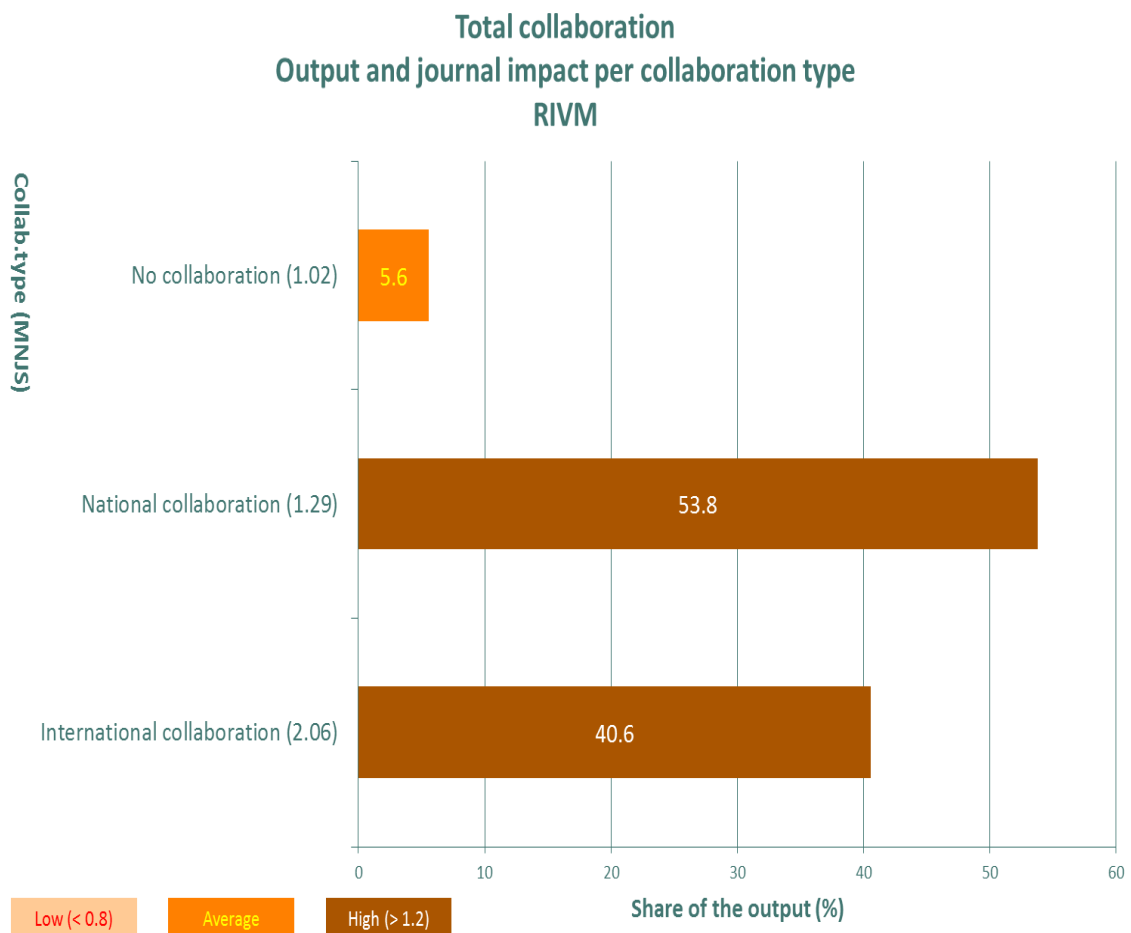
### 3.4. SOR programme collaboration.

Now that we have established the impact of the scientific fields SOR is concentrating its research output in, we would like to show to what extent these papers were a joint effort or simply a single address, RIVM only, endeavor and how they measure up in impact. This is graphically represented in figure 3.10.

As shown in this graph, the SOR programme is cooperative in nature. Very few papers are published by one affiliated RIVM address only. More than 94% of all papers are produced in collaboration, of which the largest share is in national collaboration. All collaboration types show an impact that is clearly above or on average as defined by the World Average level. International cooperation has the best impact, showing a very high impact of more than 2 times World

Average. This is a common effect for international collaboration. Papers with a single address or in national cooperation are lagging behind somewhat but national collaborations are still highly cited whereas single RIVM publications are at World Average.

Figure 3.10 Impact and share SOR per collaboration type 2011-2014/15



## 4. Main findings

CWTS compiled a report of the bibliometric performance of the researchers within the RIVM institute covering their journal articles, reviews and letters. Publications from between 2011 – 2014 were identified within the Web of Science (data source by Thomson Reuters) and analyzed on the basis of citations found up to and including 2015. The publications were produced in the context of the Strategic research programme SOR 2011-2014 of the RIVM, consisting of 7 Speerpunten.

Below a bullet point summary of the previous chapters is provided:

- Data acquisition is a crucial step in any bibliometric analysis. It determines to some extent the value and the meaning of the statistics that are calculated. The results indicate **a good overall coverage (86%) for SOR publications as a whole and on the level of the Speerpunten**. This indicates that bibliometrics calculations are reliable.
- The total of 196 SOR publications (P) have received, on average, 9.46 citations (MCS) in the period 2011-2014/15. The citation impact is almost twice World Average, with an MNCS of 1.97 (1 is average). Furthermore, 18% of SOR publications belong to the top 10% highly cited publications. SOR produces **highly visible, (inter)nationally acclaimed scientific research in internationally highly acknowledged journals**. The impact level depends on a stable citation contribution across the oeuvre.
- There are **noticeable differences** between the bibliometric impact of the publications of the **Speerpunten ranging from 15% under World Average to over 4 times World Average**. 'IRA', 'ANT', 'IDD' score a high to very high MNCS, whereas 'SVR' and 'FKA' do not perform on World Average level. This indicates that the overall result is not carried equally by the different research within the SOR themes, although we must bear in mind the limited number of publications the analyses for the Speerpunten are based upon.
- These research topics of SOR show up as a wide variety of Web of Science research categories, indicating the breadth of the SOR research, and representing the core of RIVM research. **"Infectious diseases", "Public, Environmental & Occupational Health", "Environmental sciences" and "Toxicology"** all have a high MNCS. Together the **high scoring fields contribute 50% of the production of SOR**. Some research fields representing 17% of the output are lagging behind (e.g. "Biochemistry & molecular biology" and "Genetics & heredity")
- The spearheads showing **the highest impact reflect the long term core research of RIVM** except for the Strategic vaccine research, whereas the spearhead focusing on implementation in society (FKA: from knowledge to action) may need further development (or is difficult to get published). However, indicators based on a limited volume of publications need to be viewed

with caution, even though both MNCS and PPTop10% point in the same direction (indicating robustness for this finding).

- The SOR research programme is cooperative in nature. **More than 94% of all papers are produced in collaboration, of which the largest share is in national collaboration (54%).** International cooperation has the best impact, showing a very high impact of more than 2 times World Average. This is a common effect for international collaboration.

#### Point for discussion

The scientific versus societal impact of the spearheads. In the current analysis the societal impact is not taken into account. The RIVM has evaluated other outputs including criteria that were developed by the RGO/GR. These are described in their own report (RIVM2015-0103). From this analysis there is no indication that spearheads that are highly cited are also more active in other regards, or the reverse. The fact that there is no such direct or indirect link may suggest that these type of indicators are not sufficient to cover the results of all the spearheads. Alternatively, one should realize that new research ideas need time to develop and 'pay off', and that lack of measurements doesn't mean that there is no progress!

# Appendix I. Data collection, selection and handling

## Initial database structure

The data selection was performed on the basis of the RIVM supplied raw input definitions of the publications. These were converged towards the expected data-types and fed into an algorithm geared towards matching as precisely as possible but which also allows for small common inaccuracies between the matched items.

## Bibliometric approach

The CWTS Citation Index (CI) system is used for these analyses. The core of this system is comprised of an enhanced version of Thomson Reuters Scientific/Institute of Scientific Information's (ISI) citation indexes: Web of Science (WoS) version of the Science Citation Index, SCI (indexed); Social Science Citation Index, SSCI and Arts & Humanities Citation Index, AHCI.

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 journal in WoS is assigned to one or more subject categories (scientific fields).

Each publication in WoS has a document type. In the calculation of bibliometric indicators, we only take into account publications of the document types 'article', 'review' and 'letter'. In general, these three document types cover the most significantly important scientific publications. In addition, publications in multidisciplinary journals which do not have sufficient references to WoS-covered non-multidisciplinary journals cannot be assigned to a subject category and hence are excluded from the analysis. Letters are assigned a weight of 0.25 in the analysis because of their erratic cited behavior.

The final outcome of the data selection process comprises a table in which we have the UT (unique publication identifier in the WoS) of all the publications of SOR listed within the WoS. On the basis of the UT-identifier we can collect bibliographic and bibliometric data on the papers put up for analysis.

## Coverage of publications

The first step is to determine the internal coverage for SOR publications. The internal WoS coverage is defined as the proportion of the references that point to publications covered by WoS. To gain insight



in the WoS Citation Index coverage of the publications included in the study, we determined to what extent they themselves cite WoS papers and to what extent other non-WoS documents.

The internal coverage provides insight into the citing practices of SOR and, in particular, how well WoS covered SOR publications reflect the scholarly practice at RIVM and the relevance of the WoS in that respect. This we can then use as an indication of how well WoS is geared towards providing robust indicators for analysis. The internal coverage for the institute total output publications, counted whole (except for 'letters' which are always counted as a fraction), is presented in **Error! Reference source not found..1**.

Table I.1 Internal coverage for SOR.

Publication years	publications	Internal Coverage
2011-2014	195.75	85%
2011	18.00	86%
2012	39.00	86%
2013	58.25	86%
2014	80.50	85%

Table I.2 Internal coverage for the Speerpunten within SOR.

Publication years	Speerpunt	publications	Internal Coverage
2011 - 2014	ANT	21.00	86%
2011 - 2014	FKA	12.00	75%
2011 - 2014	HEA	42.00	89%
2011 - 2014	HSL	20.25	78%
2011 - 2014	IDD	47.50	84%
2011 - 2014	IRA	26.00	79%
2011 - 2014	SVR	28.00	94%

As a rule of thumb, whenever internal coverage percentage drops below 50% it is not possible to perform robust analyses with confidence as this is an indication that the non-WoS citation environment is as important as the environment within WoS. In this case the results indicate a good overall coverage for SOR publications as a whole and on the level of the Speerpunten. Only 14% of the documents cited by the SOR articles, reviews and letters are published in sources not covered by WoS. This can include books and book chapters, conference papers, reports, patents or even certain journals, as well as articles published before 1980. At that level of coverage, a robust bibliometric analysis of the publications is warranted.

We do however have to take into account when interpreting these data that the number of publications analysed at the level of Speerpunten sometimes drops to a too low level. A level at which we do not guarantee that robust outcomes can be generated.

## Appendix II. Bibliometric indicators

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

### II.1. General matters

The analysis in this report is based on publications and citations received by those publications covered by WoS. As mentioned beforehand, only the document types 'article', 'review' and 'letter' are considered. These document types account for 71.33% of total WoS output. WoS includes other 32 distinct document types and 27 of these document types are assigned to at most 1% of all publications in WoS. The other 5 frequent document types are 'meeting abstract', 'book review', 'editorial material', 'note' and 'news item'.

The articles, reviews and letters also attract more than 96% of the total citations in WoS. Nonetheless, the indicators in the report are computed using all the citations received by the publications in the analysis, regardless the document type of the citing paper. For example, we count all the citations received by a given article in the analysis, including the citations from other articles, reviews, letters but also meeting abstracts, editorial materials, etc.

It needs to be mentioned that this approach is different from the one used in Leiden Ranking, where only articles and reviews are used in the analysis. In addition, only citations originating from articles and reviews are counted, not from other document types.

Furthermore, the present analysis uses a variable-length citation window. We therefore account for all citations, from 2005 until 2013, received by the publications included in the analysis. For publications in 2005, the citations from 2005 until 2013 are considered and for publications in 2006, the citations up to 2013 are considered, therefore spanning over a 8-year citation window. Finally, for publications in 2012, we consider their citations in 2012 and 2013. Leiden Ranking uses a variable-length citation window as well, though the period of analysis is different. For example, Leiden Ranking 2014 considers publications in the period 2009-2012 and their citations until the end of 2013.

### II.2. 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. We stress that research articles, review articles and letters are the only publication types that are taken into account. Other publication types, such as editorial material, meeting abstracts, and book reviews, are not included.

### **II.3. 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 used the full counting method whenever possible and appropriate. This means that 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. This is opposed to 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. Impact indicators calculated using full counting tend to have higher values than impact indicators calculated using fractional counting. The main advantage of full counting over fractional counting is that full counting is usually perceived as more intuitive and more easy to interpret. There is however some risk that full counting gives results in which certain scientific fields are favored over others.

#### **Un-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 PnC indicator counts the number of publications that have received no citations, and the PPnC 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.

Each journal in WoS is assigned to one or more subject categories. These subject categories can be interpreted as scientific fields. There are about 250 subject categories in WoS. Publications in multidisciplinary journals such as *Nature*, *PLoS ONE*, *Proceedings of the National Academy of Sciences*, and *Science* are individually allocated, as much as possible, to subject categories on the basis of their references. The assignment of these publications to subject categories is done proportionally to the number of references pointing to a subject category. Impact indicators are calculated taking into account this assignment of publications in multidisciplinary journals to subject categories.

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 scientific 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. As mentioned before, the field (or the fields) to which a publication belongs is determined by the WoS subject categories of the journal in which the publication has appeared.

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.

In addition to the MNCS indicator, we also have the TNCS (total normalized citation score) indicator. This indicator is calculated by summing the normalized citation scores of all publications of a research unit. The TNCS indicator equals the product of the MNCS and P indicators.

Since the MNCS indicator relies 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 small 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 PPTop10%, 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 PPTop10% 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 (i.e., the same WoS subject category) and the same publication year. The PPTop10% 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 PPTop10% indicator, this means that the actual number of top 10% publications of the unit equals the expected number. A value of 20% for the PPTop10% indicator for instance means that a unit has twice as many top 10% publications as expected. We note that in addition to the PPTop10% indicator we also have the Ptop10% indicator. This indicator equals the number of top 10% publications of a research unit. The Ptop10% indicator is obtained by multiplying the PPTop10% 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 PPTop10% 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.

### **Publications belonging to multiple fields**

As explained above, a publication may belong to multiple fields (i.e., multiple WoS subject categories). In that case, the publication is fractionally assigned to each of the fields to which it belongs and normalized impact indicators are calculated accordingly. For instance, a publication may belong to two fields. In one field the number of citations of the publication may be twice above expectation, while in the other field the number of citations may be at the expected level. The normalized citation score of the publication then equals to  $(2 + 1) / 2 = 1.5$ . Likewise, a publication may belong to two fields and may be a top 10% publication in one of these fields but not in the other. In that case, the publication is considered to be a top 10% publication with a weight of 0.5. This for instance means that the publication contributes a value of 0.5 to the Ptop10% indicator.

### **Limitations of field normalization**

It is important to emphasize that the correction for field differences that is performed by the MNCS and PPTop10% indicators is only a partial correction. As already mentioned, these indicators are based on the field definitions provided by the WoS subject categories. It is clear that, unlike these subject categories, fields in reality do not have well-defined boundaries. The boundaries of fields tend to be fuzzy, fields may be partly overlapping, and fields may consist of multiple subfields that each have their own citation characteristics. From the point of view of citation analysis, the most important shortcoming of the WoS subject categories is their heterogeneity in terms of citation characteristics. Many subject categories consist of research areas that differ substantially in their density of citations.

For instance, within a single subject category, the average number of citations per publication may be twice as large in one area compared with another. The MNCS and PPTop10% indicators do not correct for this within-subject-category heterogeneity. This can be a problem especially when using these indicators at lower levels of aggregation, for instance at the level of departments or individuals.

### **Indicators of journal impact**

We use the total and mean normalized journal score indicator, denoted by TNJS and 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 the same year and on the other hand the average number of citations of all publications published in the same field (i.e., the same WoS subject category) and the same year. The TNJS indicator is obtained by summing the normalized journal scores of all publications of a research unit, while the MNJS indicator is obtained by averaging the normalized journal scores of all publications. 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. 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.

## **II.4. Indicators of scientific co-operation**

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 III. 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. (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 III.1 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  $PP_{top10\%}$  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.<sup>1</sup> 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  $PP_{top10\%}$  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.

---

<sup>1</sup> 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.

## Appendix IV. Overview underlying data files

Supplied as accompanying Excel files:

1. Collaboration Speerpunten.xlsx
2. Collaboration Totaal.xlsx
3. Overview Speerpunten.xlsx
4. Overview total.xlsx
5. Scientific Profile.xlsx