CWTS BIBLIOMETRIC REPORT

Meaningful metrics

Bibliometric research performance of the European Neuromuscular Centre (ENMC) 2000 – 2016/17

August 2018

Universiteit Leiden

Report for the European Neuromuscular Centre (ENMC)

Project team

Erik van Wijk, *Project leader*

Thanks to:

Bert van der Wurff for covering part of the data selection

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

Table of contents

1. Introduction.

1.1. Objective of the research

The Centre for Science and Technology Studies (CWTS) of Leiden University performed a bibliometric analysis on the performance of the European Neuromuscular Centre (ENMC). The goal of the project is to gain concrete and detailed insight into the bibliometric performance of the publications derived from ENMC workshops.

In 1992, a group of European patient associations that dedicated itself to bring together leading researchers and clinicians from all over the world together founded the ENMC organization. Their mission statement proclaims that: "The mission of ENMC is to encourage and facilitate communication and collaboration in the field of neuromuscular research with the aim of improving diagnosis and prognosis, finding effective treatments and optimizing standards of care to improve the quality of life of people affected by neuromuscular disorders". Therefore, ENMC mainly supports and facilitates research without being involved in the actual research itself. Whenever the term "ENMC publications" is used in the report, readers will understand that this means, "ENMC-workshop derived publications". These include workshop reports published in the journal of Neuromuscular Disorders and indirect papers which report on scientific research and/or clinical projects that were initiated or inspired by ENMC workshops.

ENMC prides itself in a longstanding and successful history in bringing consortia together.

Today the ENMC is proud of:

1) more than 235 workshops on its résumé,
2) results from more than 70% of the workshops published in the journal of Neuromuscular Disorders (Elsevier), last 10 years even reaching a publication rate of 90%.
3) an active network of over 2500 researchers, clinicians and patients with whom ENMC is in regular contact,
4) participants from over more than 65 countries creating international cross-talk and collaboration in trials.

The present study into the bibliometric performance of ENMC 2000 – 2016 will shed a light on the impact and visibility of the scientific work resulting from this encouragement and facilitation of communication and collaboration in the field of neuromuscular research by ENMC.

The data collection used throughout was based on the Web of Science (WoS, core collection): the Science Citation Index, Social Science citation Index and the Arts and Humanities Index all owned by Clarivate Analytics. This, CWTS will refer to as Citation Index (CI). The results of the analysis performed by CWTS are presented in this report. Our report focuses on the publications between 2000 - 2016 of the ENMC and the research areas attached to them. The citation count of these publications is measured during the same time period with one additional year (2017) to allow 2016 publications to also gather citations (this is denoted as 2000 – 2016/17). The citation impact is then achieved by comparison with worldwide reference values referred to as “World Average”. We will come back to this term later in the report because there are different ways of defining “World Average” with possible direct consequences.

The study is based on a quantitative analysis of scientific articles, reviews and letters published in international journals covered by CI only. The objective of the analysis is to assess the publication activity and international impact and visibility of ENMC facilitated publications. To this end, we analyzed the publication impact profiles of the ENMC as a whole. Before presenting the analyses, CWTS introduces the bibliometric terms used within the report and their implementation concisely.

1.2. Bibliometric approach and indicators overview

CWTS calculated impact indicators for ENMC based on the in-house version of the CI database. Each publication in the CI has a document type. The most frequently occurring document types are ‘Article’, ‘Book review’, ‘Correction’, ‘Editorial material’, ‘Letter’, ‘Meeting abstract’, ‘News item’, and ‘Review’. In the calculation of bibliometric indicators for this study, CWTS only took into account publications of the document types ‘Article’, ‘Letter’ and ‘Review’. In general, these document types cover the most frequently cited publication types and are therefore referred to as ‘citable items’. The normalization of the impact indicators, which makes it possible to compare different scientific fields with different citation behavior to each other, was done based on the direct environment of the publications as defined by links you can set up using matching publication characteristics and citation links.

This approach we will address under the moniker of “publication clustering normalization”. Which allows for each publication to be indexed in its own peer environment as much as possible. How this cluster buildup is done precisely is beyond the scope of this report but a paper on the subject is available at: https://www.cwts.nl/pdf/cwts-wp-2012-006.pdf. But in short, this clustering is based on citation relations between individual publications. Through this method, collections of individual papers are assigned to specific clusters. ENMC derived workshop publications are thus assigned to these pre-defined cluster(s). These cluster(s – it can be more than one) are then used for normalization.

The indicators resulting from analyses and computations used throughout the report are listed below, grouped by dimension.

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. (PPtop1% is therefore the percentage share in the top 1% cited publications etc.)
PPnC	Impact	Proportion of uncited publications.
MNJS	Journal impact	Average normalized citation impact of a journal.
No Collaboration	Collaboration	Proportion of publications authored by a single institution.
National Collaboration	Collaboration	Proportion of publications resulted from national collaboration.
International Collaboration	Collaboration	Proportion of publications resulted from international collaboration.

A more extensive explanation of these indicators and how they are computed, can be found in Appendix I and II.

2. Data collection, selection and handling

Data acquisition is a crucial step in any bibliometric analysis, as it determines largely the value and meaning of the statistics that are calculated. This section outlines the steps that were taken to ensure robustness of the findings.

2.1. Initial data selection

The initial data for ENMC comprise the results of a specific search strategy on ENMC suggested publications within the CWTS enhanced version of the Web of Science Citation Index, core collection (CI) database. The nucleus of this information system is comprised of an enhanced version of Clarivate Analytics citation indexes: Web of Science (WoS) version of the Science Citation Index, (SCI); Social Science Citation Index, (SSCI); and Arts & Humanities Citation Index (AHCI). Enhanced means that Clarivate's CI was cleaned for the multiple names used to identify one institute, harmonized institute labels were installed in the CWTS in-house database. The CWTS CI is a more unambiguous database, which however contains the same number of 12.000 journals as Clarivate and therefore the same number of publications. CWTS always uses this in-house CI to perform their bibliometric analyses. CWTS invested a great deal of effort in unification of the main institute definition in addresses of researchers contributing to publications. Additionally we collected the ENMC related publications in the Web of Science within the timespan 1992-2017, to be on the safe side of which we used in the end only 2000 – 2016 basic data, because since 2000 cluster normalization was made feasible within the in-house CI database. We used the search terms: 'ENMC' or 'EUR* NEUROMUSC* CTR' or 'EUR* NEUROMUSC* CENTER' or 'EUR* NEUROMUSC* CENTRE' within 12 different searchable fields (abstract, title, acknowledgement, author address etc.). Except for literature reference to avoid double counting of searchable data fields in the online version of the Web of Science database. In addition, CWTS matched the publication definitions ENMC provided with the Citation Index in 24 different ways. Each different way accounting for a slight possible difference in the external (ENMC supplied) data and the internal (WoS citation Index) data.

2.2. Coverage of publications

2.2.1 Internal coverage

The first step was to determine the internal coverage for ENMC publications. The internal CI coverage is defined as the proportion of the references from a publication that point in the citations to publications covered by the CI. To gain insight in the CI coverage of the publications included in the study, CWTS determined to what extent the publications themselves cite CI papers and to what extent they cite other non-CI documents. The internal coverage provides insight into the citing practices of ENMC. It is an indicator for how well the CI database reflects the scholarly practice in ENMC publications and therefore by proxy the relevance of the CI in that respect. This was used as an indication of how well the CI is geared towards providing robust indicators for the analysis. The internal coverage for the ENMC is presented in Table 2.1. As a rule of thumb, whenever internal coverage percentage drops below 50%, CWTS cannot perform robust analyses with confidence, as this is an indication that the non-CI citation environment is as important as, or even more important than, the environment within the CI used for analysis.

Table 2.1 Internal coverage for ENMC 2000 – 2016.

Institute (Year)	Internal coverage
ENMC (2000 - 2016)	88%
ENMC (2000 - 2003)	87%
ENMC (2001 - 2004)	88%
ENMC (2002 - 2005)	86%
ENMC (2003 - 2006)	89%
ENMC (2004 - 2007)	89%
ENMC (2005 - 2008)	88%
ENMC (2006 - 2009)	90%
ENMC (2007 - 2010)	86%
ENMC (2008 - 2011)	87%
ENMC (2009 - 2012)	86%
ENMC (2010 - 2013)	89%
ENMC (2011 - 2014)	90%
ENMC (2012 - 2015)	89%
ENMC (2013 - 2016)	90%

The high levels of internal coverage shown in Table 2.1 indicate that the CI is an appropriate tool for bibliometric analysis. These results meant that CWTS was able to perform the analyses with confidence and results would be robust and meaningful indicators.

3. Analysis of main impact indicators

3.1. Overall Numbers and Indicators

To obtain a complete view on the impact of ENMC workshops it is, next to the ENMC workshop reports, also interesting to look at indirect publications which are derived from ENMC workshops. Thus, a large set of papers spontaneously published within the literature, and acknowledging "ENMC" or related search terms as used by CWTS, were included in the cluster normalization to find out what the "spin-off" is of workshops and how the network is collaborating.

This revealed an output of 98 papers of which approximately 30% were ENMC workshop reports (n= 32) and 70% indirect publications (n=66). Apparently not every Neuromuscular Disorders edition was every year included in the CWTS CI database and/or not every publication identified as review or article, which explains the smaller % of ENMC reports in this selection, as compared to the total number of published workshop reports (n=171). Nevertheless, it provides a snapshot of the impact of ENMC-workshop derived papers within this period. In this section, CWTS presents the overall statistics for the ENMC. Table 3.1 shows the total number of publications and their key bibliometric indices using publication clusters for normalization. In the calculation of all impact indicators, CWTS disregarded author self-citations (see Appendix 1.3.1)

CWTS chose to use moving average year-blocks primarily to dampen sharp fluctuations in the indicators to be able to concentrate more on trend. CWTS uses 4 years as opposed to any other number of years to on the one hand allow a substantial part of the publications to reach the height of their citation returns, which is usually at between 3 and 4 year and on the other hand still have enough year-blocks to be able to actually show development.

Table 3.1 Impact ENMC publications, 2000 – 2016/17.

Year	P	MCS	TCS	MNCS	MNJS	TNCS	PP top 10%	PP un-cited	PP self-citations	PP collab	PP int collab
2000 - 2016	98	19.95	1955	1.24	1.13	121.85	16%	6%	20%	83%	56%
2000 - 2003	16	7.81	125	0.88	0.89	14.01	6%	13%	31%	81%	44%
2001 - 2004	13	12.08	157	1.39	1.23	18.03	15%	15%	31%	85%	54%
2002 - 2005	11	5.82	64	1.02	0.68	11.20	9%	9%	32%	82%	55%
2003 - 2006	11	7.09	78	1.02	0.73	11.20	9%	9%	20%	73%	45%
2004 - 2007	10	11.10	111	1.21	0.82	12.08	10%	0%	19%	70%	60%
2005 - 2008	9	8.00	72	1.90	1.28	17.08	22%	0%	11%	67%	33%
2006 - 2009	10	11.40	114	1.54	0.96	15.43	24%	10%	11%	80%	40%
2007 - 2010	19	6.37	121	1.22	1.09	23.14	16%	16%	21%	84%	47%
2008 - 2011	20	8.85	177	1.25	1.14	25.07	15%	0%	22%	75%	40%
2009 - 2012	19	6.58	125	0.95	1.02	18.11	5%	0%	29%	79%	47%
2010 - 2013	35	6.11	214	1.24	1.21	43.38	10%	11%	25%	77%	51%
2011 - 2014	30	6.57	197	1.43	1.33	42.87	23%	7%	18%	80%	60%
2012 - 2015	43	7.56	325	1.25	1.22	53.92	14%	14%	20%	86%	63%
2013 - 2016	50	9.26	463	1.27	1.20	63.68	17%	12%	23%	88%	66%

When doing bibliometric analyses we consider the MNCS and the PP Top 10% to be the central most indicators. The PP Top 10% represents the level at which highly cited papers were published. Since highly cited publications attract attention, we dub this an indication of visibility. The MNCS we can consider as an index to the overall impact. In tandem, these two indicators reinforce each other. Relatively few very highly cited papers could potentially cause a very high MNCS, however in conjunction with a high PP Top 10% and in the absence of a soaring share of uncited papers we can estimate that the statistical dispersion is not so severe as to stand in the way of robust indicators. (See also Appendix A.1.3.4)

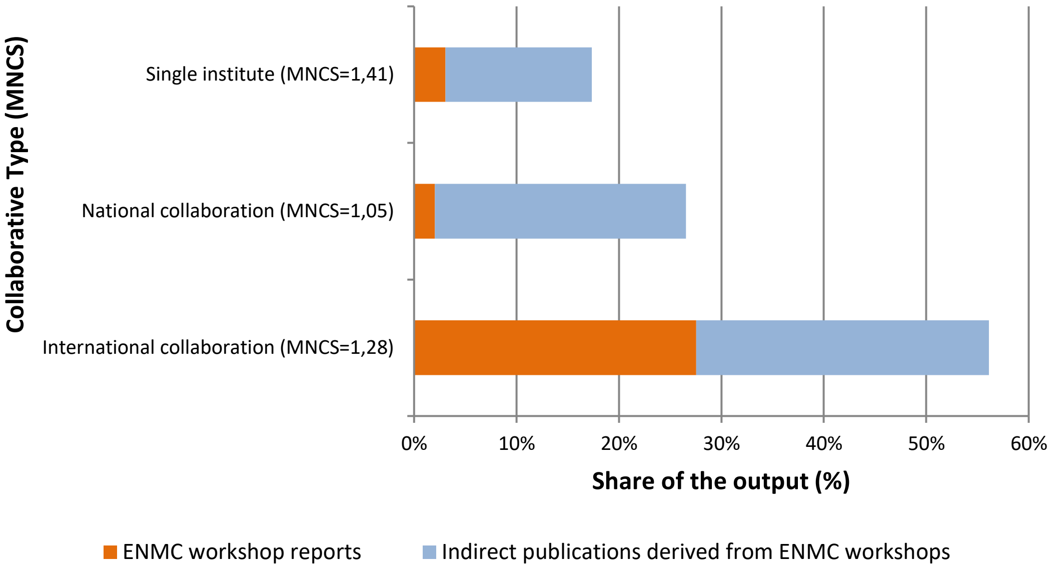
Table 3.1 shows high MNCS impact for the ENMC facilitated publications: 1.24 times world average (MNCS world average is always 1). The threshold, at which CWTS assigns the 'High Impact' label, is 1.20 times world average. In addition, the PP Top 10% representation is in most 4-year periods in line (on average PP top 10% was 16%) and evolves much in accordance with the direction of the MNCS.

At the start of the analysis year blocks, the MNJS is highly mobile, fluctuating restlessly around world average. This comes to rest in later years at a level much in accordance with the MNCS of the ENMC-workshop derived papers themselves. We may understand this as an indication of the convergence over time of the impact of the publications themselves with the media as a whole in which they are presented to the rest of the world. These media tend to be of higher quality as time evolves but not necessarily of higher quality than the ENMC-facilitated publications themselves. On the contrary, for many year-blocks the ENMC-driven publications have an impact in excess of the MNJS. However all this fluctuation comes to a halt starting at year-block 2010 – 2013 when the number of publications in the analysis increases.

3.2. Collaboration Profile

In the collaboration analyses, CWTS casts the output of the different types of collaboration analyses against their input. CWTS split the collaboration types into three types. “No Collaboration” (only one single institute address), “National collaboration” (only addresses originating from one country) and “International collaboration” (more than one country affiliated in the addresses).

Figure 3.1 Collaboration analysis for the ENMC 2000 – 2016/17



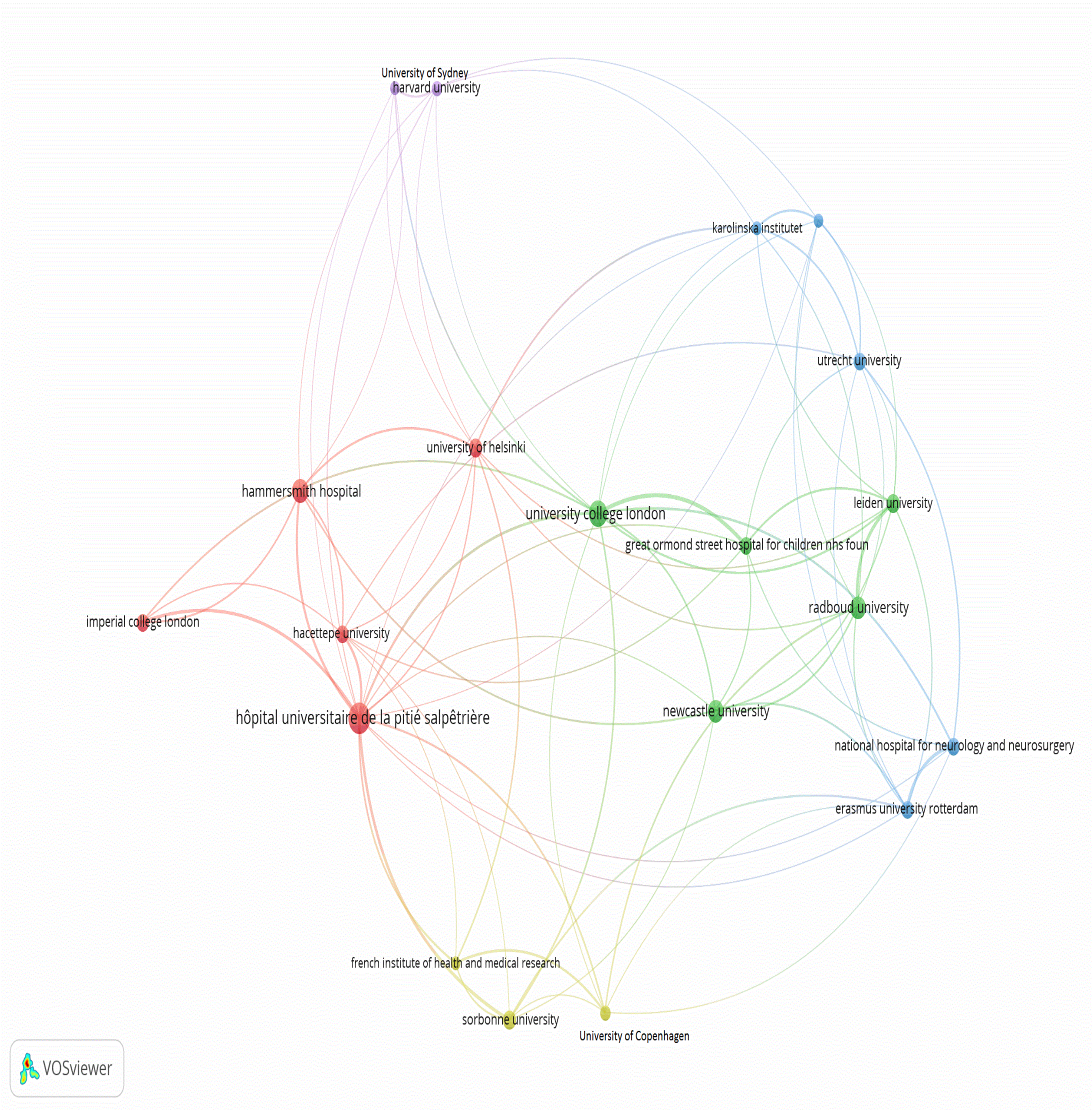
The pattern CWTS sees frequently in this type of analysis is that the international collaboration publications show a higher impact than other collaboration types. In this case, publications by a single institute/author score even higher impact and overtake the impact level of international cooperation by some 10 index points. On the other hand, 'International collaboration' scores almost 10 index points higher than the threshold for 'High Impact' at 1.2. Therefore, although the mission statement of ENMC constitutes 'to encourage and facilitate communication and collaboration in the field of neuromuscular research' and they succeed in doing that, the high share of international collaboration publications may be interpreted as testimony to that, it is not necessarily collaboration in publications that is apparently furthered by their activities. The collaboration may also lie in the inspiration that participants take home from the workshop and the contacts with other researchers in that context, which nevertheless result in a publication assigned to one author only. Also some single institute/author publications are in fact the mandatory ENMC workshop report. In the past these reports were sometimes written and published by one of the organizers, but nowadays this is being achieved by the team of workshop organizers from various countries and institutes.

Since there is no highest possible impact level defined for the MNCS but once 'High impact' is reached an even higher value than the threshold indicates relatively ever increasing levels of scientific excellence in the bibliometric definition of quality by citation indicator proxy, the difference in impact between international collaboration and no collaboration is not an absolute one.

3.3. Scientific Profile

According to their mission statement, ENMC encourages and facilitates communication and collaboration in the field of neuromuscular research. In the previous analysis on collaboration, we showed that also in scientific appraisal this is a worthwhile cause. In this chapter, we will analyze the main actors in the collaborations encouraged by ENMC. Within the cleaned in-house CI database, one name was assigned to cover multiple institutes in a city. For example, reference is made in Figure 3.2. to "Leiden University" but this entails both the Leiden University Medical Centre (LUMC) and the multiple research departments at the Leiden University.

Figure 3.2 Collaboration Network analysis ENMC 2000 – 2016



In the collaboration network analysis it is shown which institutions are more often collaborating (based on co-publications) than others. The VOSviewer tool (www.vosviewer.com) is used to discriminate clusters of collaboration. Clusters are indicated by color, the size of the circle represents the size in publication numbers. Figure 3.4 shows the most prominent institutions in each of the clusters.

- The red cluster is characterized by the central position of the Groupe Hôpital La Pitié Salpêtrière, which is collaborating with Hammersmith Hospital, University of Helsinki and Imperial College London. These institutes are publishing on ENMC topics mostly relating to general genetic and acquired neuromuscular conditions, such as Duchenne muscular dystrophy, inflammatory body myositis, CIDP, dermatomyositis, endocrine, toxic or iatrogenic myopathies, congenital myasthenic syndromes, neuropathies and mitochondrial DNA neuromuscular diseases. In the framework of the national French Plan for the study of neurodegenerative diseases (2014-2019), the site of the Pitié-Salpêtrière in Paris was selected as regional centre of excellence in the field of neurodegenerative diseases to lead translational research on neurodegenerative diseases. It hosts the Brain and Spine Institute (ICM) bringing together 650 researchers implicated in research on the neurodegenerative diseases: multiple sclerosis, Parkinson disease, Alzheimer disease, lateral amyotrophic sclerosis, movement disorders, rare neurodegenerative diseases, and transversal research aimed at studying cognition, behavior and psychiatric diseases. The central position of the Pitié-Salpêtrière in this cluster may arise from the critical mass created by the Plan.

- The green cluster represents the network relation between University College London, Newcastle University and Radboud University Nijmegen, with collaborating institutions at Leiden University and Great Ormond Hospital, publishing on topics mostly relating to congenital muscular dystrophies, LGMD, FSHD, myotonic dystrophies, Duchenne and Becker muscular dystrophy. Both Newcastle and Radboud University Nijmegen are centres of expertise in many aspects of a range of neuromuscular diseases.

- The blue cluster is smaller and brings together ErasmusMc Rotterdam, University of Utrecht and Karolinska Institute, publishing on ENMC topics mostly relating to Chronic inflammatory demyelinating polyneuropathy, Pompe's disease and Guillain-Barré syndrome.

- The yellow cluster is the smallest and stands on its own at the fringes of the network. It consists of two important centres: University of Copenhagen and Sorbonne University, publishing on topics mostly relating to **genetic** dystrophies (congenital myopathies, metabolic myopathies, congenital myasthenic syndromes, neuropathies) and **acquired** (myasthenia gravis, inflammatory, endocrine, toxic or iatrogenic myopathies, psychogenic affections).

Altogether, the collaboration network shows strong European connections based on ENMC workshop papers. The fact that Harvard University is one of the few international collaborators outside Europe is a strong sign of quality since Harvard University is among the top medical research institutions worldwide.

Of the 98 papers, 55 were international collaborations (56%). International collaboration depends on the address information in a paper. If more than one institute from different countries is associated with one or more of the authors listed, international collaboration (and hence contribution from different countries) is present.

We looked at the contribution of different countries to these 56% of international collaborations. For example, it appeared that Great Britain is involved in 31 of the international collaborative publications. The sum of the largest country contributers ranked in this table is 162, total of contributing countries sum up to 180, which is due to overlapping contributions of several countries to the same papers. The collaborations most prominently generated through the publications facilitated by ENMC are those in conjunction with Great Britain. As is shown in the table below.

Table 3.2 Country Collaboration analysis ENMC 2000 – 2016/17 (contribution of at least one institute per country)

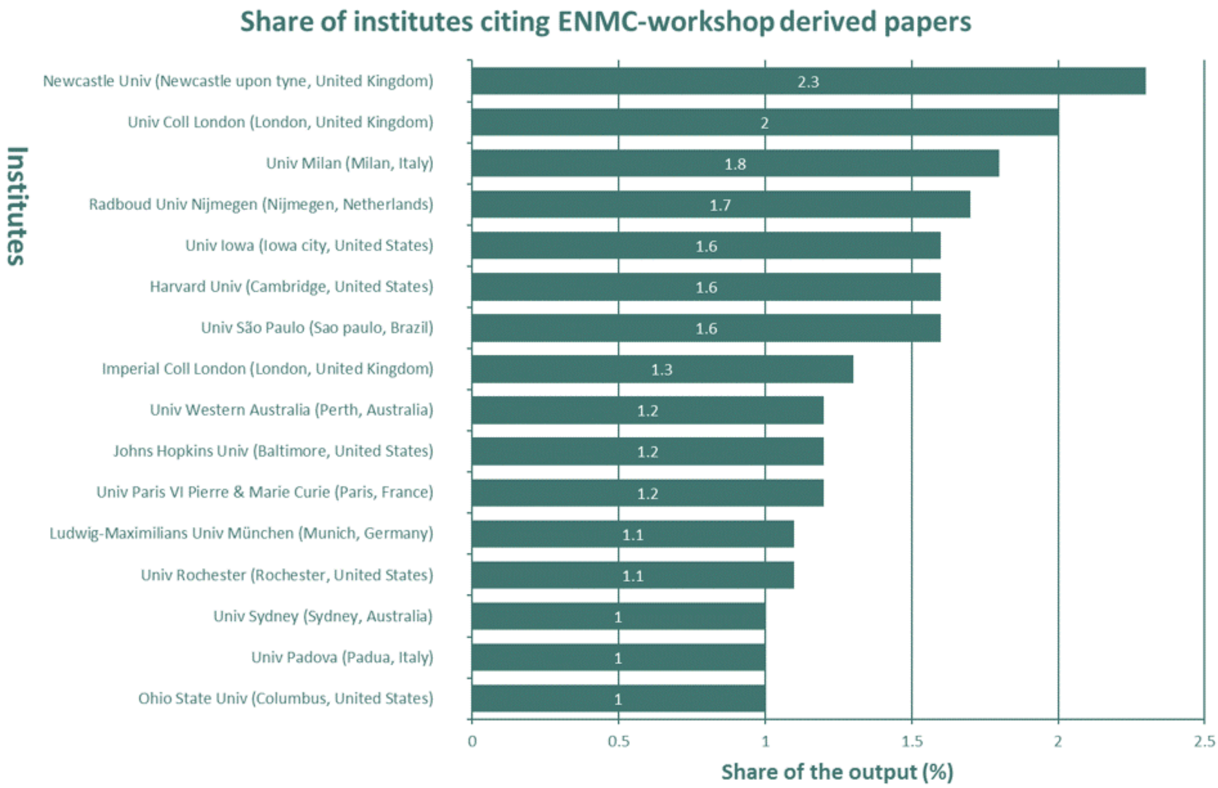
Country participation in ENMC collaborative publications											
Great Britain	France	USA	Nether-lands	Italy	Germany	Finland	Denmark	Australia	Sweden	Belgium	Canada
31	26	22	20	19	11	9	6	5	5	4	4
Number of collaborative papers											

All in all the majority of collaborations are those within Europe with a relatively small role for Canada and Australia. It is the big European countries, that are also represented in the ENMC board, who are most prominently present in the ENMC-worksop derived papers. The Scandinavian countries are less prominently recognized in this dataset.

3.4. Knowledge User Profile

CWTS analyzed the knowledge user characteristics of the ENMC-workshop derived publications on the institute level in Figure 3.3. It shows which institutes most prominently cite ENMC-workshop derived publications. By analyzing this, we identify the most significant users of the ENMC-enabled research. The institutes, ranked in descending order based on their share in the total number of citations to ENMC-derived publications, are presented in Figure 3.3.

Figure 3.3 Knowledge user analysis ENMC 2000 – 2016/17 (Share of more than 1% in total)



Most institutes which cite publications derived from ENMC workshops, are from the USA and United Kingdom. This comes as no surprise since the sheer weight of the number of publications from these Anglo Saxon countries and institutes lends credence to this scenario.

Likewise, it is no surprise that the top-4 citing institutions are those that are central in the collaborating network analysis as well. Other than that, there is a wide range of knowledge users, which is particularly active outside of Europe.

We do however find distinctly heavy weight scientific institutes among the most prominently citing institutes. Citing institute Harvard University is among the absolute most highly cited institutes in the world, and Imperial College London ranks at 33. All of these are themselves also very well cited within the context of the specific scientific research we focus on in this study.

4. Main findings

CWTS performed an impact profile analysis for the European Neuromuscular Centre (ENMC) on publications from the period 2000 - 2016 (citations counted up to 2017). The data was selected through an address search strategy on the CWTS proprietary version of the Web of Science Core Collection (CI-CWTS) database as well as an identification of CI publications based on publication definitions by the ENMC itself.

The output trend of ENMC-workshop derived publications is somewhat declining until 2008 but after that output numbers begin to rise. Impact fluctuates but is high if computed over the entire period.

CWTS presented the collaboration profile, of which three main types were identified, namely one single institute only (one address, no collaboration) publications, national and international collaborative publications. CWTS found that the normalized impact indicator was above world average for no collaboration and international collaborated publications in which the first scored the highest impact. In this case, we also witnessed a prominent role in that respect for single address publications. Prominent in impact though not so much in volume. This level of impact for that type of (non-) collaboration on the publication level, may lie in the inspiration and exchange of ideas in the workshops ENMC hosts which nevertheless result in a “non-collaborated” publication.

ENMC considers their mission: ‘to encourage and facilitate communication and collaboration in the field of neuromuscular research with the aim of improving diagnosis and prognosis, finding effective treatments and optimizing standards of care to improve the quality of life of people affected by neuromuscular disorders’. (International) cooperation is indeed abundantly present in the body of publications CWTS analyzed and a strong indicator of successfully complying with that intention. The relatively strong performance of single address publications may in that respect be a more indirect testimony to that as suggested above.

Cooperation was most prominent when British institutes were also involved. CWTS found mainly European based cooperation with some USA and Australian parties involved. The prestigious Harvard University, a top medical research institution worldwide, present in the collaboration network is a clear indication of the quality of the work furthered by the workshops of ENMC.

The knowledge presented in publications was cited and therefore used mostly by Anglo Saxon institutes.

Appendix I. Bibliometric indicators

In this appendix, CWTS describes the methods underlying the present bibliometric analysis.

A1.1. General matters

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

The articles and reviews also attract some 95% of the total citations in CI. Nonetheless, the indicators in the report are computed using all the citations received by the publications in the analysis, regardless of the document type of the citing paper. For example, CWTS counts all the citations received by a given article in the analysis, including the citations from other articles, reviews and letters, but also meeting abstracts, editorial materials, etc.

It needs to be mentioned that this approach is different from the one used in the Leiden Ranking which only counts citations originating from articles and reviews, not other document types. Furthermore, the present analysis uses a variable-length citation window. CWTS therefore accounts for all citations from 2009 until 2016 received by the publications included in the analysis. For publications 2009-2015, the citations from 2009 until 2016 are considered (effectively a 7-year maximum citation window) and for publications between 2010-2015, the citations between 2010 and 2015 are considered, therefore spanning over a 6-year maximum citation window. Finally, for the last publication year-block, 2012-2015, CWTS considered citations in 2012-2016. Obviously, this also goes for the normalization values, making it possible to compare impact results between papers with a longer and a shorter citation window.

A1.2. Output indicator

The publication output indicator, denoted by P, measures the total publication output of a research institute. It is calculated by counting the total number of publications of a research institute, including only publications covered by CI. CWTS stresses that research articles, review articles and letters are the only publication types that should be taken into account. Other publication types are not included such as editorial material, meeting abstracts, and book reviews.

A1.3. Impact indicators

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

A1.3.1 Self-citations

In the calculation of all our impact indicators, CWTS disregards author self-citations. CWTS classifies 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, CWTS ensures 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. The logic is that 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/her own work is therefore ignored.

A1.3.2 Counting method

In computing the impact indicators, CWTS used the full counting method whenever possible and appropriate. This means that publications are always fully assigned to research institutes, regardless of the collaborative nature of the authorship, e.g., single-authored, two authors from the same research institute, 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 institute. 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 easier to interpret. There is however some risk that full counting gives results in which certain scientific fields are favored over others.

A1.3.3 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 institute. 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 institute.

A1.3.4 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, CWTS uses normalized citation indicators.

Each journal in CI is assigned to one or more subject categories. These subject categories can be interpreted as scientific fields. There are about 250 subject categories in CI. 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 an institute, CWTS first calculates the normalized citation score of each publication of the institute. 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 CI 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 an institute. If an institute has a value of one for the MNCS indicator, this means that on average the actual number of citations of the publications of the institute equals the expected number of citations. In other words, on average the publications of the institute 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 an institute have been cited twice as frequently as would be expected based on their field and publication year. Please refer to Appendix II for an example of the calculation of the MNCS indicator by CWTS.

In addition to the MNCS indicator, CWTS also has the TNCS (total normalized citation score) indicator. This indicator is calculated by summing the normalized citation scores of all publications of a research institute. The TNCS indicator equals the product of the MNCS and P (Output) 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 an institute has one such publication, this is usually sufficient for a high score on the MNCS indicator, even if the other publications of the institute 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 institute.

Therefore, in addition to the MNCS indicator, CWTS uses another important impact indicator. This is PPtop10%, the proportion of the publications of a research institute that belongs to the top 10% mostly frequently cited publications in their field and publication year. For each publication of a research institute, 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 CI subject category) and the same publication year. If a research institute has a value of 10% for the PPtop10% indicator, this means that the actual number of top 10% publications of the institute equals the expected number. A value of 20% for the PPtop10% indicator for instance means that an institute has twice as many top 10% publications as expected. CWTS notes that in addition to the PPtop10% indicator CWTS also has the Ptop10% indicator. This indicator equals the number of top 10% publications of a research institute. The Ptop10% indicator is obtained by multiplying the PPtop10% indicator by the P (Output) indicator.

To assess the impact of the publications of a research institute, 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 institutes that are active in the same field.

A1.3.5 Publications belonging to multiple fields

As explained above, a publication may belong to multiple fields (i.e., multiple CI 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 is calculated as (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.

A1.3.6 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 CI 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 CI 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 Units of Analysis or individuals.

A1.3.7 Indicators of journal impact

CWTS uses the total and mean normalized journal score indicator, denoted by TNJS and MNJS, to measure the impact of the journals in which a research institute has published. For this, CWTS first calculates the normalized journal score of each publication of the institute. The normalized journal score of a publication equals the ratio of the average number of citations of all publications published in the same journal and the same year on the one hand, and on the other the average number of citations of all publications published in the same field (i.e. the same CI subject category) and the same year. The TNJS indicator is obtained by summing the normalized journal scores of all publications of a research institute, 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 an institute has a value of one for the MNJS indicator, this means that on average the institute has published in journals that are cited as frequently as would be expected based on their field. Likewise, a value of two for the MNJS indicator means that on average an institute has published in journals that are cited twice as frequently as would be expected based on their field.

A1.4. Indicators of scientific cooperation

Indicators of scientific collaboration are based on an analysis of the addresses listed in the publications produced by a research institute. CWTS first identifies publications authored by a single institution ("no collaboration"). Subsequently, CWTS identifies 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, CWTS considers 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 numbers 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 therefore has a smaller expected number of citations. It can also 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, CWTS obtained:

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

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 A2.1 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 PPtop10% indicator, CWTS uses the same example as we did for the MNCS indicator. Table A2.1 shows the bibliometric data for the five publications of the hypothetical research group that CWTS considers. 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.

Appendix III. Underlying data table list

These files contain the raw data on which the analyses in this report were based and which were sent accompanying the report:

Filename
Collaboration by Country.xlsx
Collaboration Overview.xlsx
Institute Overall Ut Clus.xlsx
User_Profile.xlsx