# Bibliometric analyses of publications from Centres of Excellence funded by the Danish National Research Foundation.

Report to the Danish Ministry of Science, Innovation and Higher Education

**Jesper W. Schneider**,

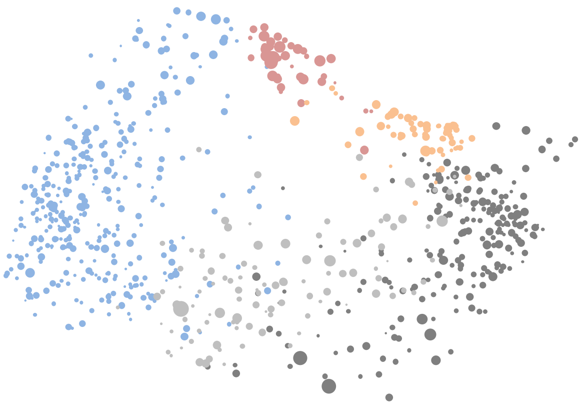
Danish Centre for Studies in Research and Research Policy, Department of Political Science and Government, Aarhus University, Denmark

*jws@cfa.au.dk*

**Rodrigo Costas**,

Centre for Science and Technology Studies, Leiden University, Netherlands

*rcostas@cwts.leidenuniv.nl*



Collaborators: Dorte Henriksen (CFA), Clara Calero-Medina and Erik van Wijk (CWTS)

# SUMMARY

Since 1993 the Danish National Research Foundation (DNRF) has funded Centres of Excellence (CoE) in Denmark. Scientific excellence is demanded from candidates, hence the selection process is hard, rejection rates high, but the eventual funding of the CoEs is very large and long-termed - up to 10 years - compared to other funding instruments. These conditions represent a unique setting for CoEs for doing outstanding research. In that respect, DNRF as a funding organization is being evaluated and the present report supplements this evaluation by investigating the overall performance of research publications coming from CoEs funded by DNRF.

To the extent that citations can somehow be considered a proxy for ‘research quality’ or rather impact in the scientific community, the goal of providing unique settings for producing outstanding research ought therefore to result in relatively high citation indicators, if indeed such outstanding research is produced. The main purpose of this report is to investigate this question.

The focus of the analysis is on DNRF and not specifically upon the individual CoEs. Consequently, the combined (and distinct) set of publications from the individual CoEs constitutes the set of publications ‘belonging’ to DNRF in this analysis. Obviously, it is somewhat arbitrary to uniquely assign publications to one specific funding institution. Publications are usually a result of several influences and funding channels. However, for convenience, we treat the DNRF-funded publications as a distinct set in this analysis (see below and Chapter 7 for a more thorough discussion and analysis of this issue). It is important to take into account that not all CoEs funded by DNRF since 1993 are included in the analysis and for those CoEs included only publications published in journals indexed by the citation database Web of Science (WoS) are analysed. If CoEs have poor publication coverage in the Web of Science, such as the humanities and computer science, or have been funded after 2009/10, they are excluded from the analysis. Furthermore, those CoEs included in the analysis may have other substantial publication activities outside Web of Science. Eventually, 66 CoEs were included in the analyses.

While the report examines a number of publication and citation indicators in order to investigate publication patterns and impact, the main focus is upon highly cited publications. In bibliometric studies, the concept of ‘excellence’ is often linked to a unit’s ability to produce highly cited publications. Citation distributions are heavily skewed, where few publications receive most citations while most publications receive few or none. Highly cited is defined as the publications cited equal to or more than a certain percentile limit of the distribution (in this report the $90^{th}$ percentile). Most of the publications at this level are considered influential and the assumption is that a unit’s proportion of publications at this level says something about the importance and excellence of this unit when it comes to influencing the scientific community. The indicator for highly cited publications is named PPtop10%, whereas the indicator for mean normalized citation scores is named MNCS.

The bibliometric analyses for DNRF-publications are done combined for all fields as well as disaggregated for a number of scientific fields and subfields. The analyses cannot say anything about the main effect of being funded by DNRF because the performance of rejected control

candidates is not investigated. However, in order to provide national and international context for the indicators calculated for the DNRF-publications, two sets of benchmark units are used for comparison: 1) the DNRF-publications' contribution to the national performance of Denmark; and 2) 10 European and American universities specifically chosen among the top-performing strata in different scientific fields in the Leiden Ranking (http://www.leidenranking.com/).

The analyses are based on elaborate, complex and labour-intensive data collection and processing procedures. The basis are publication lists from the 66 individual CoEs from which potential Web of Science journal publications have been identified, extracted, processed and eventually matched with the CWTS in-house version of the Web of Science database, based on manual work and semi-automatic matching algorithms. Likewise, publication sets for the benchmark units were established resulting in a data set comprising close to 1 million records. While voluminous, the final set of DNRF-publications used for the analyses cannot be considered complete. Some publications are missing and false positives are represented. Indeed, declaring a publication as 'belonging' to DNRF is not without some problems. Four CoEs have specifically indicated in their publication lists that some of their publications are only marginally or perhaps not at all a result of the DNRF-funding. Conversely, empirical analyses for other CoEs revealed that some publications credited to the DNRF-funding in the respective publication lists, did not acknowledge DNRF in the actual publications, though other funding agencies did receive acknowledgements. Likewise, in the batch of publication lists provided by the DNRF for the present analysis, it appeared that there were inconsistencies (e.g., one or two publication years missing) in the publication lists of at least 10 CoEs, primarily CoEs funded in the first and second period. Finally, working with a fixed time window (1993–2011) also introduces some bias as some publications which are a result of the DNRF-funding will be published after the period under investigation in this analysis. Clearly, the individual publication lists are to varying degrees inaccurate; however, as the causes of the inaccuracies are many, we have decided to treat all eligible WoS publications in the lists equally by including them all in the analysis. Investigating individual inaccuracies and subsequently setting up special inclusion criteria, as well as identifying potentially missing publications for the different CoE publication lists are too time consuming and not within the means of this analysis. Nevertheless, given that the eligible publications do appear on the publication lists of the CoEs used for reporting to the DNRF and that the focus of the analysis is the DNRF, we find it reasonable to use this selection method. Some publications are missing and some are false positives, but eventually, the large number of publications finally used for most of the analyses (in total around some 11,100) secures reliable and robust statistics, which are crucial in bibliometric analyses where distributions are heavily skewed. We have performed some sensitivity analyses and there is no indication that systematic biases, such as missed publications or false positives, have influenced the overall results.

## Results

There can be no doubt that the overall performance (all fields combined) of the DNRF-funded publications in general is impressive. Overall for the whole period analysed (1993-2011), DNRF-funded publications constitute some 7% of the Danish publications; they accumulate 10% of all Danish citations and 9% of the normalized citations. More than 20% of the DNRF-publications qualify as highly cited, i.e. among the 10% most cited publications in the database. Overall for the whole period, 14.6% of Danish publications qualify as highly cited, when excluding the DNRF-publications this indicator drops to 14.2%. Comparing the overall performance of DNRF-publications to the benchmark universities show that DNRF is ranked in the middle, very close to universities such as the University of California, San Francisco, Yale University or the Ecole Polytechnique Federale de Lausanne. Overall, the American universities have a higher performance and the European universities a lower performance, this goes for both indicators PPtop10% and MNCS.

Breaking the overall impact into annual blocks shows some fluctuations, where the PPtop10% indicator had a low-point at 16% in 2004 but three years later in 2007, the indicator peaked at 23.6%. From 2007 to 2011, the PPtop10% indicator has been at its highest levels for the whole period investigated (around 22%). This pattern is also visible in the DNRF's share of Danish publications and highly cited publications. From 1998 to 2008 DNRF's annual share of Danish publications varied around 8-9%, but from 2009 to 2011 this number has risen to around 11%. Three peaks can be seen in the annual share of highly cited publications, the first around 1998 at 10%, the second in 2001 at 12% and the third in 2010 above 14%. The first peak is achieved after a continuous rise in the share of highly cited publications since DNRF's inception year in 1993. Then there is a short drop to 8% followed swiftly by rise to 12% in 2001. Instead of stabilizing around 12% - remember the share of annual publications in this period is quite stable – the share drops considerably from 12% to 8% in 2004. Hereafter we see a rise in two steps, first to a level around 10-11% and then from 2009 to 2011, a rise to annual shares between 12% and 14%. The latter rise corresponds to a similar rise in DNRF's share of Danish publications from 2009 to 2011. Consequently, breaking down the overall performance into annual blocks shows that the annual shares of Danish publications and highly cited publications stabilizes around 1998. A new rise is seen the last three years analysed, where the influence is at its highest level for the whole period. This trend is also visible when calculating DNRF-publications' annual contribution to the Danish PPtop10% indicator. The average drop in the indicator when DNRF-publications are excluded is a half percentage point, however, there are fluctuations and in the latest years the number varies around .6 and .7 percentage points. The general trend is similar for the MNCS indicator.

Comparing the annual PPtop10% with the benchmark units, shows a trend where in 1997 the DNRF-publications are ranked in the middle around 21%, below the American and above the European universities. Then a period with fluctuations and eventually a drop to the low point in 2004 at 16%, where the DNRF-publication set are ranked below all the benchmark universities except University of Leeds. The rise in the subsequent period brings the DNRF-set above the performance of Yale University and University of California, San Francisco in the peak year of 2007 at 23.6%. The subsequent stabilizing period around 22% again ranks the DNRF-publications in the middle between the American and European universities, very close to Yale University, Cambridge University and Ecole Polytechnique Federale de Lausanne. The annual PPtop10% indicator for DNRF-publications fluctuate considerably more than most of the benchmark units. This is a normal bibliometric phenomenon, which is due to the smaller publication set of DNRF. Most benchmark units are considerably larger than the DNRF-publication set, the most comparable in relation to size is Ecole Polytechnique Federale de Lausanne and their publication numbers are between 1.5 and 2 times larger than the DNRF-set. Several factors can cause these annual fluctuations and changes in publication output are one of them. In the present analysis, the annual output of DNRF-publications is obviously dependent on the active CoEs in a given year. Hence, annual fluctuations is also related to the different funding inception years (1993/4, 1997/8 etc.) and the flux this produces in CoE activities.

The DNRF-publication set clearly performs as well as or slightly better than some of the highest-performing universities in Europe. That the DNRF-publication set can be considered as a high-performing ‘unit’ is also visible from characteristics such as its distribution of publications to the highest percentile classes in the citation distribution and the proportion of its publications that are not cited. In both respects the DNRF-set show well-known patterns of high-performing units, i.e., more publications than expected among the most highly cited and a low proportion of uncited papers, in this case around 5% which in fact is similar to Harvard University, tough Harvard’s publication output is 16 times larger than the DNRF-set.

The performance of the DNRF-publications has also been analysed in relation to different subject fields. Two classification levels of fields have been applied: main scientific fields and subfields, the latter being a disaggregation of the former. Fields are defined by a number of Web of Science journals and publications are assigned to fields according to the journal they are published in. One category ‘multidisciplinary journals’ is obviously not a field but constitute a number of journals such as *Science, Nature, PLOS One* and *PNAS* (notice, for normalized indicators, most publications in these journals are re-classified into fields based on their reference behaviour).

Not surprisingly, the large majority of DNRF-publications can be categorised as belonging to the ‘medical and life sciences’ and ‘natural sciences’. From the inception year in 1993 until 2007 more publications in the ‘medical and life sciences’ are found compared to the ‘natural sciences’, but it is noticeable that this trend has reversed since 2007 (with the ‘natural sciences’ now producing more publications than the ‘medical and life sciences’).

The contribution of DNRF-publications to the Danish PPtop10% in the different main fields in four different time periods shows that when removing these publications only miniscule drops in the national indicator, between .1 and .2 percentage points, is visible for ‘medical and life sciences, whereas drops are somewhat larger for the ‘natural sciences’ between .4 and 1.0 percentage points in the four different periods. But most noticeable, the drops are largest in the ‘multidisciplinary journals’ category, where drops in the indicator values are between 2.2 and 5.9 percentage points. Obviously, the relative number of publications, i.e. the DNRF-set compared to the Danish output, in fields is a determining factor when it comes to the size of these drops. The number of publications and highly cited publications in the ‘multidisciplinary journals’ category is relatively low compared to the other fields, however, among the highly cited Danish publications in this category, the DNRF-set contributes on average with one-third of these highly cited publications in the periods investigated.

Breaking the main fields into subfields reveals a more nuanced pattern. The trend in the four year-period blocks is that generally all subfields accrue more and more publications over the years. In the last period, 2008 to 2011, all subfields (including 'multidisciplinary journals') have more than 150 publications and five of these subfields have more than 1000 publications, these are 'physics and materials science', 'life sciences and biology', 'chemistry', 'clinical medicine' and 'biomedicine'. The contribution to the Danish PPtop10% indicator at the level of subfields shows more variation over the period. With varying degrees of contribution to the national indicator in different periods, subfields such 'astronomy and astrophysics', 'physics and materials science', 'chemistry' and 'geosciences' show sustained considerable contributions. In later periods, 'life science and biology' does the same.

When comparing the DNRF-publications to the benchmark universities on the main field level for the whole period combined we generally see the same pattern as in the analyses of all fields combined, i.e., that the DNRF-set is placed somewhat in the middle between the American and European universities. There are, however, three noteworthy differences at the main field level. First, in the 'multidisciplinary journal' category, the DNRF-publications perform at the absolute highest-level equal to MIT and Stanford University and slightly above Harvard University. This is the case for both the PPtop10% and MNCS indicators. Second, the PPtop10% indicator in the 'natural sciences' ranks the DNRF-set among the American universities just above the University of California, San Francisco and the gap to the European universities is noticeable. Third, compared to the benchmark universities in the 'medical and life sciences', the DNRF-set ranks the lowest except for Leeds University according to the PPtop10% indicator and one rank higher (slightly above Imperial College London) according to the MNCS indicator. It should also be mentioned that in this field, the DNRF-set of publications is not the smallest compared to the benchmark units. This is, the Ecole Polytechnique Federale de Lausanne albeit their lower production is markedly higher than that of the DNRF-set in terms of impact.

However, it should be emphasized that in all fields, the DNRF-publication set always perform above the international level (i.e. values of MNCS higher than 1 and higher than 10% for PPtop10%) and for all fields except the 'social and behavioural sciences' well above the national Danish level.

Analysing the indicators in relation to fields and subfields gives a more nuanced impression of the varying performances of the DNRF-publications. A similar analysis was done on the level of the 66 individual CoEs. As expected, considerable variation in output and performance among the individual CoEs were found, some performing extremely well, but also some 12-13% of the CoEs performs at or slightly below the international level. Obviously, some CoEs included in this report are still active and publishing, whereas funding to others has expired. An analysis grouping CoEs according to their granting year revealed that the median level of performance in fact were higher for CoEs funded in the later periods 2005 and 2007, that is CoEs which most probably are still active, compared to earlier granting periods where funding has expired.

Publications with international collaboration receive on average more citations than other publications. However, American universities have a very strong collaboration with each other (only one in three publications have authors from international institutions) and combined with the general bias in the Web of Science towards Anglo-American journals, this produces a somewhat different impact pattern where national collaboration often is the highest performing. DNRF-publications reveal a third pattern. The expected pattern would be relatively highest performance for publications with international collaboration and relatively lowest impact for publications with no collaboration (i.e., authors coming from the same institution). But this is not the case. While publications with international collaboration do have the highest impact around 22% for the PPtop10% for the whole period (some 57.5% of all DNRF-publications have at least one author from an international institution), publications with no collaboration have a PPtop10% value of 20% and publications with national collaboration have a value of 17%. Somehow, publications from members of CoEs affiliated to the same Danish research institution have substantial international visibility. Among the benchmark universities, only MIT, Harvard and Stanford have PPtop10% indicator values above 20% for the publications with no collaboration.

The previous analyses show that DNRF-publications belonging to the 'multidisciplinary journal' category performed extremely well. While some journals in this category are obviously considered highly prestigious, a more general analysis of publication and performance in 'high prestige' journals was also carried out. For this analysis, 'high prestige' journals are defined as those journals that have published 30% or more of the top 10% highly cited publications in their fields in a year. In other words, high prestige journals are journals that in a given year-field combination was able to attract and publish almost 1/3 of the 10% most cited publications of their fields. This is a straight-forward definition which is free of the limitations of the Journal Impact Factor. For the whole period combined, the DNRF-publication set has the largest share of publications in 'high prestige' journals compared to the other European benchmark universities; again American benchmark universities are above. The annual share of DNRF-publications in 'high-prestige' journals varies between 10 to 15% usually somewhere in-between but the pattern is slightly decreasing towards the end of the period. Except for the last three years, the DNRF-publications have the highest annual share among the European benchmark universities. In the last three years the share is almost equal around 13% except for Leeds University at 9%. Also, the annual drop in the share of Danish publications in the 'high prestige' journals varies between .3 and 1.0 percentage points when the DNRF-publications are excluded. An analysis using the MNJS (mean field normalized journal score) corroborates these results.  DNRF-publications are generally published in journals that have a higher field normalized impact as compared to all the other European benchmarks. Again, the DNRF set of publications are ranked in-between the American and the European benchmark universities, although in the most recent years, the differences between these two main blocks in terms of journal visibility has been reduced.

A final reflection concerning the issue of treating a publication as 'belonging' to a funding institution is necessary - obviously this is somehow problematic. Publications initially 'belong' to authors who are affiliated and employed by research institutions. These are normally the units of analysis in bibliometric analyses. However, researchers and research groups, on top of their base funding, often also receive external funding, such as CoE-funding by DNRF. In order to approximate in what respect a publication can be credited as ‘belonging’ to a funding institution and thus by implication be a product of that particular funding, an exploratory analysis of the funding acknowledgements present in Web of Science publications from 2008 onwards was carried out for the DNRF-publications. The purpose was restricted to analysing DNRF and other Danish funding agencies. Some 2700 publications were analysed, 29% only acknowledged the DNRF, 47% acknowledge the DNRF and at least one other Danish funding agency, and interestingly 24% had acknowledgements to other Danish funding agencies, albeit not the DNRF even though the publications are affiliated to a CoE. The results are only suggestive, but they stress the challenges of ‘ownerships’ of publications and emphasizes that publications can be a result of many influences and several funding organizations. One successful funding often leads to another; this is the well-known phenomenon of preferential attachment, also popularly known as the “Matthew effect”.

Succeeding the main analyses reported above, two supplementary analyses were commissioned and carried out. The approach and methodology of the two analyses are characterized by being novel – designed for two specific tasks – and they are explorative.

The first analysis tries to estimate the ability of CoEs – funded by DNRF – to ‘recruit’ new scientists which are able to publish at least one highly cited publication within a short time period after their first detectable publication in WoS. We estimate an annual ‘recruitment’ rate for DNRF and compare this to five benchmark countries including Denmark. We find the methodology valid and reliable and the results are clear and robust. Hence, the estimated ‘recruitment’ rate for DNRF is genuine. Overall, DNRF has the highest ‘recruitment’ rate of the six units analysed. The ‘recruitment’ rate is approximately 50% for most years analysed; however, there is a minor drop from 1999 to 2003, where the rate drops to approximately 45%. In this short period, the ‘recruitment’ rate for Switzerland is on par or slightly above the rate of DNRF. Hence, in general one in two scientists affiliated to CoEs funded by DNRF have been associated with at least one highly cited publication within three years of their first identified publication in WoS. Also, for the whole period, 14.3% of the new scientists identified for Denmark are associated with CoEs funded by DNRF and 17.4% of the ‘successful’ new scientists in Denmark are associated with DNRF.

The purpose of the second supplementary analysis is to explore DNRF’s involvement in potential ‘breakthrough’ research. This is done by use of refined citation analyses and large-scale clustering of journal articles from 1993 to 2011 in WoS to detect potential ‘breakthrough’ papers. ‘Breakthrough’ research is obviously a challenging concept to define, operationalize and detect. We want to emphasize that our approach is a modest one, certainly with limitations, but also an interesting one in as much as we seek to detect potential ‘breakthrough’ papers among an exclusive set of the most highly cited papers in WoS in a carefully constructed network of relevant research fields. We define a ‘breakthrough’ paper as: ‘a highly cited paper, with an important spread over its own field(s) and also other fields of science, and it must be a paper that is not a mere follower of other highly cited publication(s) but that it has a genuine relevance on its own’. Three distinct ‘breakthrough’ detection-approaches are explored and compared to a sample of eight CoEs which are considered to have produced potential ‘breakthrough’ research.

Overall the results are to a large extent consistent with the sample of examples provided to us and the empirical results show some robust and well-known characteristics. In all three approaches, the distribution of potential ‘breakthrough’ papers is heavily skewed over CoEs, thus most detected potential ‘breakthrough’ papers are associated with few CoEs, and for many CoEs, one or no potential ‘breakthrough’ papers are detected, i.e. the latter to varying degrees depending on the restrictedness of the specific the detection-approach used. Approach 1 is the most restrictive; here 21,670 potential ‘breakthrough’ papers are detected in WoS from 1993 to 2011. Of these, 0.15% or 32 papers come from the set of DNRF-supported publications. Notice, the set of DNRF-supported publications constitutes 0.07% of all publications analysed, thus there is an overrepresentation of potential ‘breakthrough’ papers from the set of DNRF-supported publications. It is noteworthy that 12 of these potential breakthrough’ papers can be categorized as research in bioinformatics and eight papers as research in nanoscience and that these two research topics (and their three CoEs) are the most visible in all three approaches.

In the second approach, which is the least restrictive, 179,349 publications from 1993-2011 have been detected as potential ‘breakthroughs’, of these 241 come from the set of DNRF-supported publications and this corresponds to 0.13% of all potential ‘breakthrough’ papers defined by this approach. Besides bioinformatics and nanoscience, one more highly visible research topic appears with this approach: register-based epidemiological research. Also visible is research topics such as catalysis, metal structures, as well as muscle and sensory-motor research. The third approach is based on approach 2 but more restrictive as it focuses on knowledge diffusion; here a total of 59,617 articles are considered potential ‘breakthroughs’ according to this approach. Of these, 0.16% or 97 papers come from the set of DNRF-publications and this corresponds to approximately 1% of the DNRF-publications. It is very important to emphasise that we have only focused on the most significant signals in the DNRF-publication set. More detailed analyses will indeed show potential ‘breakthrough’ papers for a considerable number of the CoEs. While it is not entirely surprising that the same three to four research areas and CoEs recur in the results of all three approaches, after all the approaches are variations over the same idea, it is noticeable that these areas come out so strong. Also, the three clearly highest ranked CoEs in the PPtop10% impact analysis in the main report are also the three most prominent in this ‘breakthrough’ analysis (bioinformatics and nanoscience). In that respect, the ‘breakthrough’ analysis substantiates the main findings in the report.

The conclusion of this report is that the DNRF-publications overall perform at a very high-level comparable to the absolute highest-performing universities in Europe and often slightly better. The DNRF-publications contributes notably to the overall Danish impact given its relative size. However, there are annual fluctuations and marked variations in performance between fields and subfields and between individual CoEs. Especially, noteworthy is the performance in the category ‘multidisciplinary journals’, here the DNRF-publications performs at the same level as the highest ranking universities in the world.

# 1. INTRODUCTION

The present study provides insight into the publication output and international citation impact of journal publications from Danish Centers of Excellence (CoE) funded by the Danish National Research Foundation (DNRF). The unit of analysis is DNRF, hence the combined set of publications from the different CoEs constitutes the publications 'belonging' to DNRF (for convenience we name these DNRF-publications). Consequently, the individual CoEs are not the main focus in the analyses; however we do provide some insight into the individual differences in publication behaviour and citation performance between CoEs in order to qualify the overall results. Publication and citation performances for DNRF-publications are analysed combined for all fields as well as disaggregated for a number of main fields and subfields. Citation impact is compared to worldwide reference values, however, in order to give more context to the results, the DNRF-publications are also compared to a number of benchmark units, which include Denmark (i.e., national performance with and without DNRF-publications), as well as ten specifically chosen European and American universities. The time period covered is 1993-2011 for publications, including an extra year for their citation period so as to arrive at robust impact scores. This period allows most units to produce a number of publications sufficient for statistical analysis.

Citations do not measure research quality *per se*; however, it is assumed that citation impact *may* reflect a dimension of research quality. Impact refers to usefulness (Martin & Irvine, 1983), i.e., scientists cite their colleagues' work when it is useful for their argument, which means that the cited work has had a certain impact on the citing author's work. Objections have been raised against this premise, but it is widely assumed that with a sufficient number of publications and especially when focusing on the most highly cited publications, which in the skewed distributional universe of citations attracts some 60% of all citations, sound quantitative evaluations can be made; however, given the indistinct or partial relation to research quality, citation impact analyses should always be subjected to peer evaluations or better inform the latter.

The study is based on a quantitative analysis of scientific articles published in journals and serials indexed in the Web of Science (WoS) versions of the Science Citation Index and the Social Science Citation Index (SSCI); here the CWTS database containing these records, as well as enhanced citation data is briefly indicated as CI. Using bibliometric techniques, the present study analyses the publication output and citation impact of publications from DNRF. Notice, the study only focuses upon publications from journals indexed in the WoS. As the international journal coverage of WoS greatly varies between different fields and languages, and as some fields do not have journal publication as their main scientific communication channel, some CoEs are excluded from the analysis and some CoEs included in the analysis will have a substantial number of scientific publications excluded from the analyses. Hence, in the present analysis, publication output cannot be seen as an indicator of total publication activity for individual CoEs.

DNRF-publications are identified through a labour-intensive process where publication lists from eligible CoEs were manually scrutinized in order to detect potential journal publications. These were extracted, processed and subsequently matched with records in CI-WoS. Consequently, the total number of DNRF-publications should be treated cautiously as there will be false positives as well as missing publications. Nevertheless, on the aggregate analysis level of DNRF, results are robust given given the number of publications involved.

The structure of this report is as follows: Chapter 2 gives a general description of the methodology (including the data collection processes that were carried out) and terminology used and an overview of the bibliometric indicators that were calculated in the study. Chapter 3 describes the overall performance indicators of the DNRF, Denmark and the benchmark universities, including the main bibliometric results for all universities included in the study (main results, trend analysis and thematic profile analysis). Chapter 4 focuses on the performance of the DNRF and the different benchmark units across main fields of science. Chapter 5 presents the trend analysis of the publication and impact of the different units under study. Chapter 6 presents results regarding the individual Centres of Excellence involved in the study. Chapter 7 outlines the results regarding the analysis of funding acknowledgment data regarding the DNRF. Chapter 8 presents the collaboration analysis results. In Chapter 9, the analysis of the publication in high impact journals is presented. Chapter 10 presents two extra analyses commissioned subsequent to the main analyses reported in the previous chapters; the two analyses concern DNRF's ability to produce new scientist that publishes highly cited papers and an analysis that explores the possibilities of bibliometric methods to detect 'breakthrough research' papers. Both methods and results for these analyses are reported in this chapter. Chapter 11 discusses some of the caveats and limitations of the study, and finally Chapter 12 presents the main conclusions of this study.

# 2. TERMINOLOGY AND METHODS

Here, we outline the methods underlying the bibliometric analyses presented in this report[1].

## Database Structure

At CWTS, we calculate our indicators based on our in-house version of the Web of Science (WoS) database of Thomson Reuters. 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'. We notice that our in-house version of the WoS database includes a number of improvements over the original WoS database. Most importantly, compared to Thompson Reuters' WoS, our database uses a more advanced citation matching algorithm and an extensive system for address unification. The database also supports a hierarchically organized field classification system on top of the WoS 'subject categories' constructed by Thompson Reuters.

To determine the appropriateness of our indicators for assessing a particular research unit, the internal WoS coverage of the unit is examined. The internal WoS coverage of a unit is defined as the proportion of the references in its oeuvre that points to publications (also) covered by WoS. The lower the internal WoS coverage of an entity's output, the more careful one should be in the interpretation of our indicators. The rest of this chapter provides an in-depth discussion of the main bibliometric indicators that we use in this report.

**Table 2.1. Overview of standard CWTS bibliometric indicators.**

| Indicator | Dimension | Definition |
|---|---|---|
| P | Output | Total number of publications of a unit. |
| int_cov | Output | Internal coverage. Proxy of oeuvre being covered by Web of Science. Measured by the proportion of cited references in the oeuvre linking to other WoS publications. |
| MCS | Impact | Mean number of citations of the publications of a unit (self-citations not included). |
| TCS | Overall | Total number of citations. |
| MNCS | Impact | Mean normalized number of citations of the publications of a unit (self-citations not included). |
| TNCS | Overall | Total average normalized number of citations. |
| MNJS | Journal impact | Mean normalized citation score of the journals in which a research unit has published. |
| p top 10% | Overall | Number of publications belonging to the top 10% highly cited publications in the database. |
| pp top 10% | Impact | Proportion of papers that belong to the top10% highly cited publications in the database. |
| P unicited | Overall | Number of uncited publications. |
| pp uncited | Overall | Proportion of papers uncited |
| Prop self cits | Overall | Proportion of self-citations |

[1] We refer to Moed (2005) for a general introduction to the use of bibliometrics and citation analysis for research evaluation.

| | | |
|---|---|---|
| pp collab | Collaboration | Percentage inter-institutional collaborative publications |
| pp int collab | Collaboration | Percentage international collaborative publications |

## Indicators of output

To measure the total publication output of a unit, we use a very simple indicator. This is the number of publications, denoted by P. This indicator is calculated by counting the total number of publications of a research unit. In the calculation of the total number of publications, articles and reviews have a weight of one, while letters have a weight of 0.25. This is due to the fact that they are not considered publications on the same level as full articles and results in decimal values in the counting of the P indicator. In the present analysis we use full counting of publications, thus a unit is credited with a full P if at least one author from the unit under investigation is mentioned in the author by line of a publication.

## Indicators of impact

A number of indicators are available for measuring the average scientific impact of the publications of a unit. These indicators are all based on the idea of counting the number of times the publications of a unit have been cited. Citations can be counted using either a fixed-length citation window or a variable-length citation window. In the case of a fixed-length citation window, only citations received within a fixed time period (e.g., three years) after the appearance of a publication are counted. In the case of a variable-length citation window, all citations received by a publication up to a fixed point in time are counted, which means that older publications have a longer citation window than more recent publications. An advantage of a variable-length window over a fixed-length window is that a variable-length window usually yields higher citation counts, which may be expected to lead to more reliable impact measurements. In this analysis we have mainly applied variable-length citation windows.

In the calculation of our impact indicators, self-citations are disregard. We classify a citation as a self-citation if the citing publication and the cited publication have at least one author name (i.e., last name and initials) in common. We disregard self-citations because they have a somewhat different nature than ordinary citations. Many self-citations are given for good reasons, in particular to indicate how different publications of a researcher build on each other. However, sometimes self-citations can serve as a mechanism for self-promotion rather than as a mechanism for indicating relevant related work. This is why we consider it preferable to exclude self-citations from the calculation of our impact indicators. By disregarding self-citations, the sensitivity of our impact indicators to manipulation is reduced. Disregarding self-citations means that our impact indicators focus on measuring the impact of a work on other members of the scientific community.

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, Proceedings of the National Academy of Sciences, and Science are, if possible, individually re-assigned to subject fields on the basis of their references. The reassignment is done proportionally to the number of references pointing to a ‘subject category’. It is important to highlight that the overall impact indicators are calculated based on this assignment. When we disaggregate the analyses to major and subfields, it is necessary to keep the multidisciplinary journal category due to reassignment process not being exhaustive and the problem of duplicates. In Chapter 4, we describe a somewhat different field delineation with major fields based on the aggregation of WoS 'subject categories'.

Each publication in WoS has a document type. The most frequently occurring document types are article, book review, correction, editorial material, letter, meeting abstract, news item, and review. In the calculation of bibliometric indicators, we only take into account publications of the citable document type article, letter, and review. Publications of other document types usually do not make a significant scientific contribution.

The most straightforward impact indicator is the mean citation score, denoted by MCS. This indicator simply equals the average number of citations of the publications of a unit. Only citations within the relevant citation window are counted, and self-citations are excluded. Also, only citations to publications of the document types: article, letter, and review are taken into account. In the calculation of the average number of citations per publication, articles and reviews have a weight of one while letters have a weight of 0.25.

A major shortcoming of the MCS indicator is that it cannot be used to make comparisons between scientific fields. This is because different fields have very different citation characteristics. For instance, using a three-year fixed-length citation window, the average number of citations of a publication of the document type article equals 2.0 in mathematics and 19.6 in cell biology. So it clearly makes no sense to make comparisons between these two fields using the MCS indicator. Furthermore, when a variable-length citation window is used, the MCS indicator also cannot be used to make comparisons between publications of different ages. In the case of a variable-length citation window, the MCS indicator favours older publications over more recent ones because older publications tend to have higher citation counts.

The mean normalized citation score, denoted by MNCS, provides a more sophisticated alternative to the MCS indicator. The MNCS indicator is similar to MCS except that it performs a normalization that aims to correct for differences in citation characteristics between publications from different scientific fields, between publications of different ages (in the case of a variable-length citation window), and between publications of different document types (i.e., article, letter, and review[2]). 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 in WoS that belong to the same field and that have the same publication year and the same document type. The field (or the fields) to which a publication belongs is determined by the WoS 'subject categories' of the journal

[2] We note that the distinction between the different document types is sometimes based on somewhat arbitrary criteria. This is especially the case for the distinction between the document types *article* and *review*. One of the main criteria used by WoS to distinguish between these two document types is the number of references of a publication. In general, a publication with fewer than 100 references is classified as *article* while a publication with at least 100 references is classified as *review*. It is clear that this criterion does not yield a very accurate distinction between ordinary articles and review articles.

in which the publication has appeared. The MNCS indicator is obtained by averaging the normalized citation scores of all publications of a unit. Similar to the MCS indicator, letters have a weight of 0.25 in the calculation of the average, while articles and reviews have a weight of one. If a unit has an MNCS indicator of one, 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 on par with similar publications in terms of field, publication year, and document type. 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, publication year, and document type. We refer to Waltman, Van Eck, Van Leeuwen, Visser, and Van Raan (2011a; 2001b) for more details on the MNCS indicator.

To illustrate the calculation of the MNCS indicator, we consider a hypothetical research group that has only five publications. Table 2.2 provides some bibliometric data for these five publications. For each publication, the table shows the 'subject category', 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.) All publications are of the document type: article. Citations have been counted using a variable-length citation window.

**Table 2.2: Bibliometric data for the publications of a hypothetical research group.**

| Publication | Subject category | 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 |

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 and the same document type. Publication 5 also belongs to the same field and has the same document type. 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 and the same document type. 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 1, 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 fields, publication years, and document type.

Perhaps the most important impact indicator is the proportion of highly cited publications, in this analysis the proportion of publications among the 10% most cited in the database, denoted PPtop10%. For each publication of a research unit, this indicator determines whether, based on its number of citations, the publication belongs to the top 10% of all WoS publications in the same field (i.e., the same WoS 'subject category') and the same publication year and of the same document type. The PPtop10% indicator equals the proportion of the publications of a research unit that belong to the top 10%. Analogous to the MCS and MNCS indicators, letters are given less weight than articles and reviews in the calculation of the PPtop10% indicator. If a research unit has a PPtop10% indicator of 10%, this means that the actual number of top 10% publications of the unit equals the expected number, which is 10% (i.e. the citation distribution of a unit's publications is expected to follow the reference standard, thus it is expected that its 10% most highly cited publications are also among the 10% most highly cited in the global distribution, if the PPtop is set to 10%). A PPtop10% indicator of, for instance, 20% means that a group has twice as many top 10% publications as expected. Of course, the choice to focus on top 10% publications is somewhat arbitrary. Instead of the PPtop10% indicator, we can also calculate for instance a PPtop1%, PPtop5%, or PPtop20% indicator. In this study, however, we use the PPtop10% indicator. On the one hand this indicator has a clear focus on high impact publications, while on the other hand the indicator is more stable than for instance the PPtop1% indicator.

To illustrate the calculation of the PPtop10% indicator, we use the same example as we did for the MNCS indicator. Table 2.2 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 and of the same document type.[3] 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:

$$\text{PP}_{\text{top 10\%}} = \frac{2}{5} = 0.4 = 40\%$$

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

To assess the impact of the publications of a unit, our general recommendation is to rely on a combination of the MNCS indicator and the PPtop10% indicator. The MCS indicator does not

[3] 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, a publication year, and a document type we end up with exactly 10% top 10% publications.

correct for field differences and should therefore be used only for comparisons of groups that are active in the same field. An important weakness of the MNCS indicator is its strong sensitivity to publications with a very large number of citations. If a unit has one very highly cited publication, this is usually sufficient for a high score on the MNCS indicator, even if the other publications of the group 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 unit. The PPtop10% indicator is much less sensitive to publications with a very large number of citations, and therefore does not suffer from the same problem as the MNCS indicator. A disadvantage of the PPtop10% indicator is the artificial dichotomy it creates between publications that belong to the top 10% and publications that do not belong to the top 10%. A publication whose number of citations is just below the top 10% threshold does not contribute to the PPtop10% indicator, while a publication with one or two additional citations does contribute to the indicator. Because the MNCS indicator and the PPtop10% indicator have more or less opposite strengths and weaknesses, the indicators are strongly complementary to each other. This is why we usually recommend taking into account both indicators when assessing the impact of a unit's publications.

Within this report however we focus mainly upon the PPtop10% because we are interested in 'excellence' and the proportion of highly cited publications is considered to be a better indicator of this than average based citation indicators such as MNCS; further we also use PPtop10% because of its robustness, i.e., few missing publications will not influence the indicator, this provides us with a steady and sturdy measure that is very well geared towards benchmarking over a longer period of time. To complement the PPtop10% indicator, we do report the corresponding MNCS when deemed necessary. The data on which we base our findings and all the other indicators are supplied in full in the appendices for further reference material.

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, the field definitions on which these indicators are based on 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 characteristics. From the point of view of citation analysis, the most important shortcoming of the WoS subject categories seems to be 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 50% larger in one research area than in 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.

Finally, we use the mean normalized journal score indicator, denoted MNJS, to measure the impact of the journals in which a unit has published. To calculate the MNJS indicator for a unit, 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 on the other hand the average number of citations of all publications published in the same field (i.e., the same WoS ‘subject category’). Only publications in the same year and of the same document type are considered. The MNJS indicator is obtained by averaging the normalized journal scores of all publications of a unit. Analogous to the other impact indicators, letters are given less weight than articles and reviews in the calculation. The MNJS indicator is closely related to the MNCS indicator. The only 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 an MNJS indicator of one, this means that on average the group has published in journals that are cited equally frequently as would be expected based on their field. An MNJS indicator of, for instance, two means that on average a unit has published in journals that are cited twice as frequently as would be expected based on their field citation activity. The use of Thomson Reuters’ Journal Impact Factors and the MNJS indicator seems to be similar in the sense that in both cases publications are assessed based on the journal in which they have appeared. However, among several deficiencies, Journal Impact Factors have the important disadvantage that they do not correct for differences in citation characteristics between scientific fields. Because of this disadvantage, impact factors should not be used to make comparisons between fields. The MNJS indicator, on the other hand, does to a large extent correct for field differences. When between-field comparisons need to be made, the use of the MNJS indicator can therefore be expected to yield significantly more accurate journal impact measurements than the use of impact factors.

## Data collection

The data collection for this project has been the most challenging element of the analysis because of the lack of an existing database on the publications supported by the DNRF. Instead, DNRF collected and delivered printed and electronic publication lists from their CoEs; these lists were previously generated by the individual CoEs for their reporting of activities to DNRF. The main challenge was to identify and extract publications eligible for analysis in WoS from these lists, a task that required both substantial manual and algorithmic work. The quality and form of the publication lists varies substantially. Many different publication types and activities are included in the lists and publications ‘prepared’, ‘submitted’, ‘accepted’ or ‘in press’ are often also mentioned. For this reason, and a general lack of standardization, unidentifiable publications and duplicates are to be expected, and it is also highly unlikely that the publication lists can be considered exhaustive.

Notice only publications from journals covered by WoS will be included in the analysis and the number of publications for a CoE is not exhaustive and *only* reflects its publication activity in WoS-journals and therefore cannot be considered an indicator of its general publication activity. It is very important to emphasise that in individual publication lists, some eligible publications will be missing and false positives are present. Indeed, declaring a publication as ‘belonging’ to DNRF is not without some problems. Four CoEs have specifically indicated in their publication lists that some of their publications are only marginally or perhaps not at all a result of the DNRF-funding. Conversely, empirical analyses for other CoEs revealed that some publications credited to the DNRF-funding in the respective publication lists, actually did not acknowledge DNRF in the actual publications, though other funding agencies did receive acknowledgements. Likewise, in the batch of publication lists provided by the DNRF, it appeared that there were inconsistencies (e.g., one or two publication years missing) in the publication lists of at least 10 CoEs, primarily CoEs funded in the first and second period. Clearly, the individual publication lists are to varying degrees inaccurate; however, as the causes of the inaccuracies are many, we have decided to treat all eligible WoS publications in the lists equally by including them all in the analysis. Investigating individual inaccuracies and subsequently setting up special inclusion criteria, as well as identifying potentially missing publications for the different CoE publication lists are too time consuming and not within the means of this analysis. Nevertheless, given that the eligible publications do appear on the publication lists of the CoEs used for reporting to the DNRF and that the focus of the analysis is the DNRF, we find it reasonable to use this selection method. Some publications are missing and some are false positives, but eventually, the large number of publications finally used for most of the analyses secures reliable and robust statistics, which are crucial in bibliometric analyses where distributions are heavily skewed. Subsequent sensitivity analyses showed no indication that systematic bias, such as missed publications or false positives, have influenced the overall results. Considering this situation, the main approach for the data collection has been carried out as follows:

Step 1. Identification of eligible CoEs. In principle, the analysis includes all CoEs from the first granting year in 1993 until now. Initially all centres were eligible for inclusion, however, a number of CoEs are excluded due to the general poor coverage of their subject fields in WoS. All CoEs from the humanities and computer sciences were excluded up front. Further, social science CoEs were excluded if the journal coverage in their respective publication lists were below 50%. Also, CoEs beginning their activities in 2011 are not included in the analysis due to the short period of publication activity and the even shorter citation window for potential publications from these CoEs. Hence, 77 CoE were candidates for inclusion, 11 were excluded and 66 CoEs were included in the analysis.

Step 2. Processing and formatting the raw publication data. This step concerns identification of candidate journal publications in the publication lists. The candidate publications were extracted, either manually or automatic depending on the format of the publication lists, and processed in an excel sheet, where key bibliographic data are listed such as first author surname, first author initials, title of publication, journal name, volume number, issue number, first and last page number, publication year and identification of whether the publication is published or not according to the status mentioned in the publication list. The purpose with these standardized excel sheets is to use them as input for the matching process with CWTS's CI-WoS database. The quality of the publication lists varies considerable. Publication lists from recent CoEs are either pdf-files of word or excel files, or simply word or excel files, and they are very essay to clean and process. However, publication lists from older CoEs are typically scanned pdf-files from annual reports and the scanned quality is very poor. It was therefore not possible to copy and paste the information from these scanned pdf-files without a considerable amount of error. Hence, these publication lists required a substantial amount of time to process and it was unfortunately not possible within the given time frame to ensure that all errors were corrected. A two pronged strategy was chosen for these poor quality publication lists. Some of them were of so poor quality that we chose to search them manually in WoS and extract their WoS-accession number (UT-code) needed for the bibliometric analyses. This was very time consuming and only 6 publication lists (aka 6 CoEs) went through this process. The other publication list was treated as if they were comparable to those of good quality and was processed the same way in the excel sheet (knowing that there will be a considerable number of spelling errors in there). The sheets were then used for regular algorithmic matching, but since there is lot of errors in these sheets; non-matches was expected. These non-matches was subsequently dealt with through the manual search procedure mentioned above in order to establish whether the publication is in fact indexed in WoS and if so, extracted with the accession number. In total, 16292 records have been collected.

Step 3. Matching of records. These 16292 records have been matched against the WoS in order to find and link the DNRF records to actual WoS-publications in the database. This matching has been done in two sub-steps. Step 3.1. Automatic matching: Algorithmically the two databases have been matched, normally looking for matches considering the combination of different bibliographic elements, these including the first author, the publication date, journal of publications, volume, issue, pages, doi, etc. Step 3.2. Manual matching of missing publications: All records that could not be matched in the database based on step 3.1 have been manually checked in order to find new matched records in WoS. As a result of step 3.1 and 3.2, 14398 original records were matched with the Web of Science and obtained a UT code (i.e. the internal WoS accession number). This means that we have found more than 88% of the original records in our database, a very good matching rate. The 1894 unmatched publications corresponded normally to publications not written in English, in journals not covered in the Web of Science or to records without enough information to uniquely be identified in the Web of Science.

Supplementary data collection in relation to analysis of funding acknowledgements. From 2008 WoS has automatically parsed and indexed funding acknowledgements in journal articles (for an overview on this new piece of information in WoS, cf. Costas & van Leeuwen, 2012; Costas & Yegros-Yegros, 2013). We use this information for a specific analysis in this report and have, on top of the already identified DNRF-publications from 2008 onwards, also collected publications from WoS where DNRF are mentioned in the Funding Acknowledgment (FA) section of the publications from 2008 onwards. Duplicates were removed, however, this also means that some publications were not included in the lists provided by the CoEs but they could be linked to the DNRF based on the FA information. In any case, these DNRF publications were excluded from the output and citation impact analyses given the fact that they could not be linked to any CoE (see Table 2.4 below).

Classification schemes for subject fields. In some of the analyses the overall publication output and citation performance are broken down to main fields and in a few analyses further down to subfields. Initially we rely on the NOWT-high classification scheme developed by CWTS for the "Science and Technology Indicators" report in the Netherlands. As indicated above, arts and humanities are excluded from this analysis; this leaves us with following main fields: Engineering Sciences, Medical and Life Sciences, Multidisciplinary Journals, Natural Sciences and Social and Behavioral Sciences. The NOWT-high classification scheme is basically a mapping of WoS' 252 journal subject categories; see Appendix 1 for the basic mapping. However, for a number of analyses where DNRF are compared to other Danish research publications, we found it more suitable to break down the five main fields further into 12 subfields: Multidisciplinary Journals, Physics & Materials Science, Engineering, Clinical Medicine, Chemistry, Geosciences, Agricultural & Environmental Sciences, Astronomy & Astrophysics, Life Sciences & Biology, Mathematics, Statistics & Computer Science, Social & Behavioural Sciences and Biomedicine. This breakdown was done by using more elaborate versions of the NOWT-scheme in order to merge categories and map journals.

Benchmarking units. In scientometric analyses (as in other empirical comparisons) it is clearly most appropriate to compare like with like, e.g., countries with countries or institutions with institutions. The challenge with DNRF-publications, and eventually the CoEs is that no obvious or easily identifiable comparable unit of analysis exists. CoEs are highly privileged research groups that potentially experience short and long term benefits of various kinds due to lavish funding, including the benefits of the Matthew effect. CoE status is prestigious and strived for and it is likely that applicants in the final rounds of peer reviews are close to equal performance at that point in time. The ideal benchmark for an investigation of the effect of DNRF-funding would therefore be rejected applicants. Such an analysis would be able to measure the benefits or marginal effects (impact) of funding, compared to not being funded. Such data are not available. This essentially means that we compare the DNRF-publication performance with units that are conceptually different and thus not directly equivalent. This needs to be taken into consideration when comparing units in the final analysis. A considerable effort is needed if highly comparable benchmark units should be identified and selected publication sets created and this is not within the scope of the current analysis. Consequently we use two benchmark approaches: 1) to analyse the influence of DNRF-publications upon the national Danish citation impact and 2) select 10 universities from Europe and USA from the Leiden Ranking and compare DNRF-publications to these, overall and in main research fields.

## Selected CoEs and benchmarking units

A total of 66 CoEs were included in the analysis. CoEs are initially funded for a five-year period after which they can receive a further five-year funding period depending on an evaluation. In general, after 10 years DNRF-funding expired.

The Ministry of Science, Innovation and Higher Education has chosen the following benchmark universities: Harvard University, Stanford University, Yale University, University of California San Francisco, Massachusetts Institute of Technology (MIT), University of Cambridge, University College London, Imperial College London, University of Leeds and École Polytechnique Fédérale de Lausanne. The criteria for selection were five universities from USA and five from Europe, one top 10 ranking university, respectively from USA and Europe were chosen from each of the five main subject categories in the 2013 Leiden Ranking. Due to data collection limitations the analysis of the benchmark units is limited to the period 1997-2011.

Double occurrences of papers are excluded within each unit of analysis. So, one paper, assigned to two or more different research units, is counted only once on a higher level of aggregation.

## Affiliation data in the Web of Science

Based on the affiliation addresses of authors contained in the publications, publications were collected for the benchmark units. All relevant publications from the database years 1997 - 2011 were extracted from the CWTS in-house Web of Science database. In this sense, it is important to take into account that papers in non-WoS source journals are not included in this study for any of the units of analysis (i.e. this is a Web of Science based study). It is also important to take into account that a few journals are only partially processed for the CI. In summary, only publications processed by the WoS were included.

Articles were assigned to the benchmark universities according to the affiliations of authors, as included in the corporate address field. First, we selected articles containing the name of a university (and its major departments) explicitly in the address. Secondly, from affiliated teaching hospitals additional articles were selected that were published by authors who showed strong collaboration links with a university, as its name appeared in the address lists of at least half of their papers. In this way, for instance, a part of the papers containing the address Addenbrookes Hospital was assigned to Cambridge University (see Waltman et al., 2012, for further information). In general terms we follow the same methodology for the address data collection as presented in the Leiden Ranking[4].

## Fields/Subfields selected

The scientific 'subject categories' are a composition of several individual WoS 'subject categories' (i.e. subfields). Publications that belong to more than one subject category within the same broader field are counted only once. The description and configuration of these fields and subfields (i.e. WoS to NOWT subject categories) is presented in Appendix I.

## Coverage of Web of Science publications in references of DNRF-publications

To gain insight in the CI coverage of the DNRF, we studied the references of the WoS publications included in the present study. To this end, references in the set of DNRF-publications (1993-2011) were matched with our extended CI-WoS database (1980-2011). In this way, it was possible to estimate the reliance upon WoS publications in the set of DNRF-publications by determining to what extent they themselves cite such WoS publications. Due to the dimension of our database, we could only trace references dated between 1980 and 2011. Results are displayed in Table 2.3. Whenever internal coverage (coverage) level drops below 50% the results of the subsequent bibliometric analyses should be treated with extreme caution. Citation traffic within our database should be more – preferably much more important than that outside our database to allow for meaningful indicators.

**Table 2.3: Number of publications (articles and reviews) and internal coverage of the different units.**

| Units | P (1993-2011) | Coverage |
|---|---|---|
| DNRF total (1) | 11621 | 86.8% |

[4] cf. http://www.leidenranking.com/ and http://www.leidenranking.com/methodology/datacollection

| | | |
|---|---|---|
| DNRF analysis (2) | 11103 | 86.7% |
| Denmark | 166429 | 80.3% |
| Denmark (No DNRF) (3) | 154808 | 79.7% |
| **Units** | **P (1997-2011)** | **Coverage** |
| HARVARD UNIV | 169131 | 85.9% |
| STANFORD UNIV | 75209 | 80.8% |
| YALE UNIV | 57091 | 82.7% |
| UNIV CALIF SAN FRANCISCO | 55750 | 88.2% |
| MIT | 55009 | 79.5% |
| UNIV CAMBRIDGE | 74304 | 80.4% |
| UNIV COLL LONDON | 69257 | 83.3% |
| IMPERIAL COLL LONDON | 61249 | 84.1% |
| UNIV LEEDS | 32071 | 76.4% |
| ECOLE POLYTECN FEDERALE L | 21781 | 79.8% |

(1) This refers to all the publications that can be attributed to the DNRF in Denmark by any of our data collection approaches (i.e. including also publications detected through their Funding Acknowledgements but without a CoE). It is important to keep in mind that the absolute number of DNRF publications (i.e. including DNRF total + DNRF publications outside Denmark + DNRF publications only detected through their FA) amounted to 12485 in the period 1993-2011, but this value has not been used in any of the subsequent analysis as this is not comparable to any of the other benchmark units.

(2) This is a more restrictive selection of publications for the DNRF. In this case we only include publications that belong to one CoE and are in Denmark. This is in order to simplify our analytical approach and facilitate the comparability of the CoEs with the other benchmark units.

(3) This refers to the Danish publications excluding any DNRF publication (i.e. excluding the 'DNRF total' publications). The reason for this decision of excluding also those publications only with a FA to the DNRF (but not linked to any CoE) is that we could expect that these only FA-detected publications will probably have a similar citation pattern as those of the DNRF (cfr. Costas & Yegros-Yegros, 2013, detected the same pattern for the Austrian FWF publications). For this reason, we decided to focus on the Danish publications that do not have any relationship with the DNRF at all (i.e. excluding also those publications with a FA to the DNRF).

Some of the results of total output values reported in Table 2.3 may be slightly different with the ones presented in the following tables of results on citation impact. The reason of these minor discrepancies is due to the fact that for our citation analyses we reclassify the publications form the WoS subject category 'Multidisciplinary Sciences' to other WoS 'subject categories', in order to make the citation analysis more accurate and robust. This reclassification is based on the analysis of the cited references of the original publications. However, sometimes this reclassification is not possible for a marginal portion of the publications (e.g. when they do not carry references or they are not WoS covered publications). In those rare cases these publications must be excluded from the citation analysis and that's the reason why the values of the indicator P are sometimes slightly smaller compared to those in Table 2.3.

# 3. OVERALL PERFORMANCE

In this chapter, we present main results of the units involved in the analysis. Table 3.1 presents the main bibliometric indicators (including impact indicators) regarding the Danish units (i.e. DNRF and Denmark as a whole and with and without DNRF-publications).

**Table 3.1 Main bibliometric indicators for the Danish units (1993-2011).**

| | P | MCS | PP (uncited) | MNCS | MNJS | PP (top 10%) | TCS | P (uncited) | TNCS | P (top 10%) |
|---|---|---|---|---|---|---|---|---|---|---|
| DNRF analysis | 11102 | 28.7 | 5.9% | 1.86 | 1.5 | 20.3% | 318978 | 651 | 20690.5 | 2254 |
| Denmark | 166399 | 20.1 | 10.4% | 1.37 | 1.19 | 14.6% | 3346203 | 17372 | 228366.9 | 24287 |
| Denmark (No DNRF) | 154779 | 19.5 | 10.8% | 1.34 | 1.16 | 14.2% | 3020999 | 16643 | 206738 | 21923 |

Table 3.1 shows how the DNRF-funded publications represent around 7% of all Danish publications covered in WoS in the period 1993-2011. Compared to Denmark, DNRF-publications have a higher raw impact (MCS) and a higher field normalized impact (MNCS). Both Denmark and DNRF-publications are above the international impact reference level for the MNCS (i.e., equal to one). But clearly the impact of the DNRF-publications is 86% higher than the international level, while the same value for the Danish publications (including DNRF) is 37% above the international level. A similar pattern is observed for the impact of the publications in relation to the journals, where the DNRF-publications are published in journals with higher actual impact compared to the same value for Denmark (i.e. MNJS of 1.5 for DNRF and of 1.19 for Denmark).

Regarding the production of highly cited publications (i.e. PPtop10%), more than 20% of the DNRF-publications can be qualified as such, while 14.6% of all the Danish publications are highly cited.

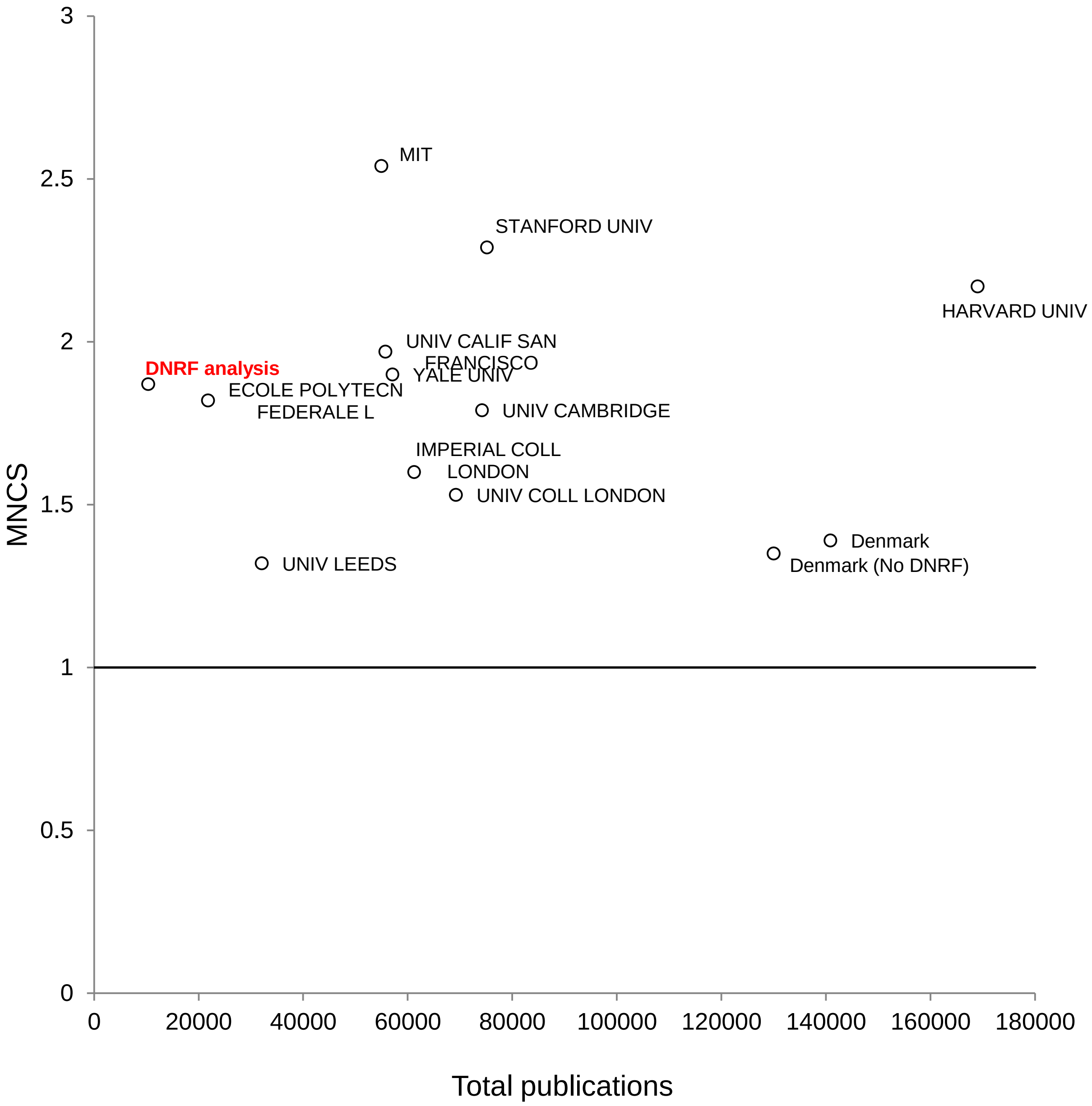
Table 3.2 presents the same indicators as in Table 3.1, but in this case referring only to the period 1997-2011 in order to be able to compare with the benchmark institutions. In essence, the same patterns are also observed for this shorter period; the set of DNRF-publications maintains its preeminent position within the Danish scientific landscape.

**Table 3.2: Main bibliometric indicators for the Danish units (1997-2011).**

| | P | MCS | PP (uncited) | MNCS | MNJS | PP (top 10%) | TCS | P(uncited) | TNCS | P(top 10%) |
|---|---|---|---|---|---|---|---|---|---|---|
| DNRF analysis | 10342 | 27.97 | 6.2% | 1.87 | 1,51 | 20.3% | 289241 | 640 | 19361.2 | 2101 |
| Denmark | 140871 | 18.89 | 11.0% | 1.39 | 1,21 | 14.9% | 2661193 | 15431 | 196321.9 | 20930 |
| Denmark (No DNRF) | 130012 | 18.2 | 11.3% | 1.35 | 1,18 | 14.4% | 2365731 | 14713 | 176022.6 | 18718 |

Figure 3.1 below shows a general overview of the performance of the set of DNRF-publications compared to all the benchmark units (both Denmark and international institutions) in terms of impact and output.

**Figure 3.1. DNRF and benchmark units comparative analysis (MNCS: Mean Normalized Citation Score; P: Publications); notice publication window 1997-2011.**




DNRF-publications are in terms of impact at a very similar level compared to benchmarks such as the University of California San Francisco, Yale University or the Ecole Polytechnique Federale de Lausanne. The set of DNRF-publications outperforms some important benchmarks such as the University of Cambridge, Imperial College London or the University College London.

As Table 3.3 and Figure 3.1 shows, the strongest benchmarks in terms of impact are MIT, Stanford University and Harvard University, all of them with MNCS higher than 2 and with more than 25% of their publications among the 10% most cited in the database (PPtop10%).

**Table 3.3: Main bibliometric indicators for the benchmark units (1997-2011).**

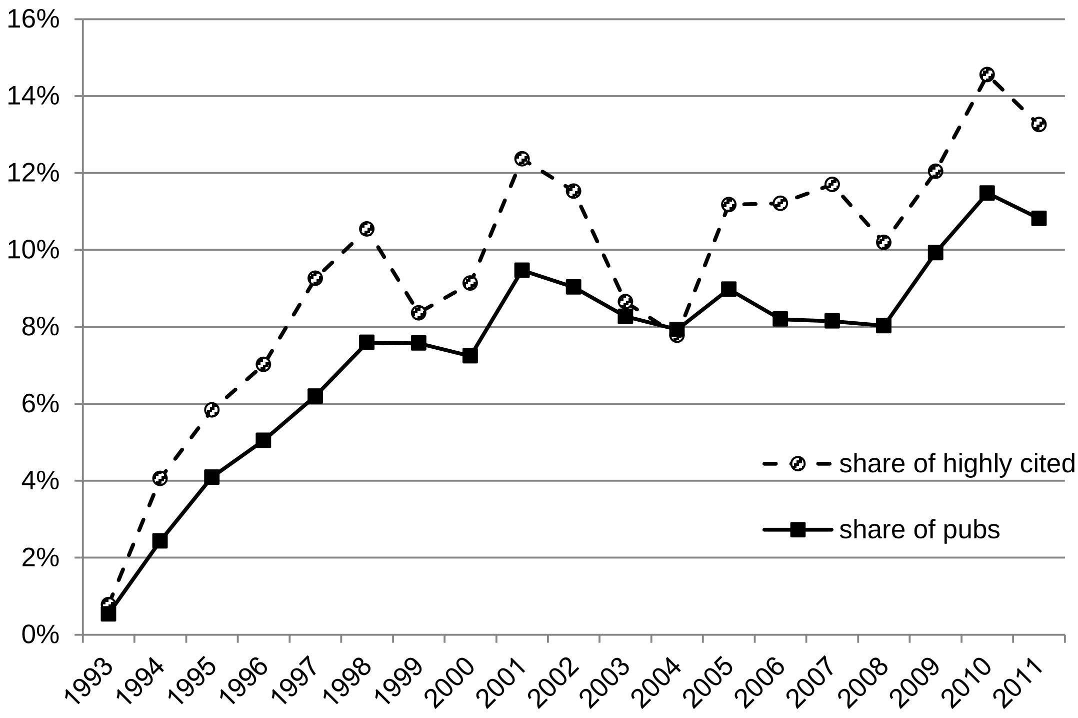
| | P | MCS | PP (uncited) | MNCS | MNJS | PP(top 10%) | TCS | P(uncited) | TNCS | P(top 10%) |
|---|---|---|---|---|---|---|---|---|---|---|
| HARVARD UNIV | 169013 | 3.,84 | 6.6% | 2.17 | 1.76 | 25.2% | 6394638 | 11200 | 366013.6 | 42529 |
| STANFORD UNIV | 75152 | 32.97 | 8.5% | 2.29 | 1.71 | 25.3% | 2477723 | 6360 | 171728.9 | 18977 |
| YALE UNIV | 57071 | 31.2 | 8.8% | 1.9 | 1.64 | 21.8% | 1780800 | 5026 | 108271.2 | 12446 |
| UNIV CALIF SAN FRANCISCO | 55735 | 37.39 | 5.2% | 1.97 | 1.65 | 23.4% | 2084022 | 2901 | 109725.9 | 13051 |
| MIT | 54939 | 34.06 | 9.6% | 2.54 | 1.84 | 27.7% | 1871362 | 5246 | 139475.7 | 15227 |
| UNIV CAMBRIDGE | 74207 | 24.95 | 11.4% | 1.79 | 1.43 | 19.1% | 1851545 | 8459 | 133196.2 | 14153 |
| UNIV COLL LONDON | 69212 | 24.59 | 9.2% | 1.53 | 1.37 | 17.2% | 1701793 | 6359 | 105573.2 | 11921 |
| IMPERIAL COLL LONDON | 61223 | 22.84 | 9.7% | 1.6 | 1.36 | 17.4% | 1398559 | 5912 | 98219.8 | 10621 |
| UNIV LEEDS | 32061 | 16.94 | 12.3% | 1.32 | 1.23 | 14.0% | 543163 | 3939 | 42349.5 | 4499 |
| ECOLE POLYTECN FEDERALE L | 21777 | 18.42 | 13.7% | 1.82 | 1.41 | 19.7% | 401052 | 2972 | 39661.4 | 4290 |

## Development in DNRFs share of highly cited Danish publications

In the following we analyse the annual contribution of DNRF-publications to the general Danish research output and citation impact.

Figure 3.2 below shows the annual share of DNRF-supported publications in *the set of all* Danish publications in a given year, as well as the annual share of DNRF-publications in *the set of highly cited* Danish publications in a given year.

**Figure 3.2: Developments in the annual share of DNRF-publications among Danish and highly cited Danish publications. The threshold for highly cited is the 90th percentile (top 10%).**



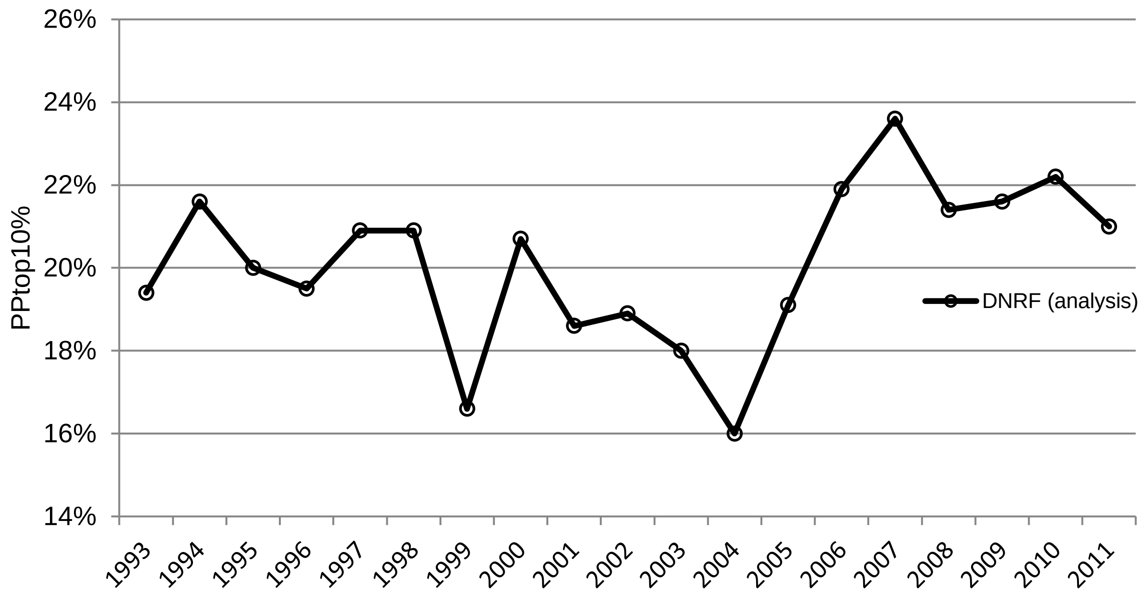
In 1998, six years after the inception, the share of DNRF-publications peaks for the first time at 7.5% of Danish WoS-publications. From 1998 to 2000, the share is quite stable around 7%, and then it rises in 2001 to 9.5%, but then falls for three consecutive years to 8% in 2004. Hereafter the share is again quite stable around 8% until 2009 and 2010, where in the latter year the total annual shares peak for the whole period at 11.5%, and this number drops in 2011 to 10.8%, though this is still the second largest number in the 19-year period analysed. For a longer period in the 2000s the share fluctuated around 8-9%, whereas in recent years this number has risen above 10% of annual Danish paper output in journals indexed in WoS.

Turning to the share of DNRF-publications among the 10% most highly cited Danish publications, we can see that in general the curve shows a higher share of highly cited papers compared to the annual output, hence more DNRF-publications than expected seems to become highly cited, however, there are some interesting drops, especially in 2003 and 2004, where the generally higher than expected number of highly cited papers levels out. It is clear from Figure 3.2, that in the subsequent analyses, we can expect to see an overrepresentation (to the expected 10%) of DNRF-publications among highly cited Danish papers.

## PPtop 10 % for Danish publications with and without DNRF publications

Figure 3.3 shows the development in the annual PPtop10% indicator for DNRF-publications.

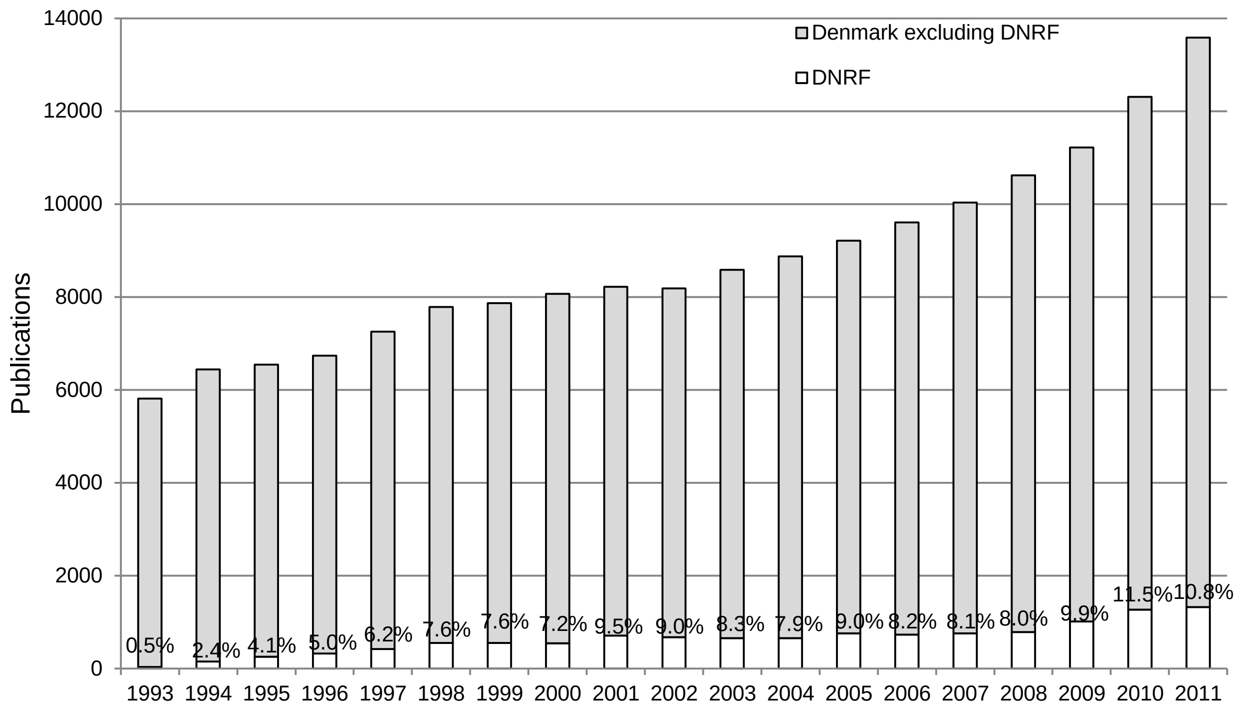
**Figure 3.3: Trend analysis of PP top 10% indicator for DNRF production compared to Denmark (overall and without DNRF publications).**




1993 is the inception year, but we see how DNRF-publications from the beginning have a substantially higher representation than expected of highly cited publications. Starting around 19% (close to twice the expected), the indicator fluxes around 20%, with lows around 16% in 1999 and 2004, and peeking in 2007 with 23.6% (136% more papers among the 10% most cited in the database). The rise from 16% in 2004 to 23.6% in 2007 is noticeable – a rise of 7.6 percentage points in merely three years. Indicator values for PPtop10% above 20% are impressive, but when we compare with benchmark units we have to consider volume as well. Smaller publication numbers makes the indicator more vulnerable to fluctuations and as units become larger, indicators tend to be more robust but also for most units to settle closer to the reference value, except for the really high performing units.

Figure 3.4 below shows the annual development in Danish WoS-publications included in the analysis from 1993 to 2011. Likewise at the bottom of the annual bars are shown the share of DNRF-publications among the set of Danish publications.

**Figure 3.4: Annual number of Danish WoS-publications included in the analysis; the share of DNRF-publications is provided at the bottom of the bars.**

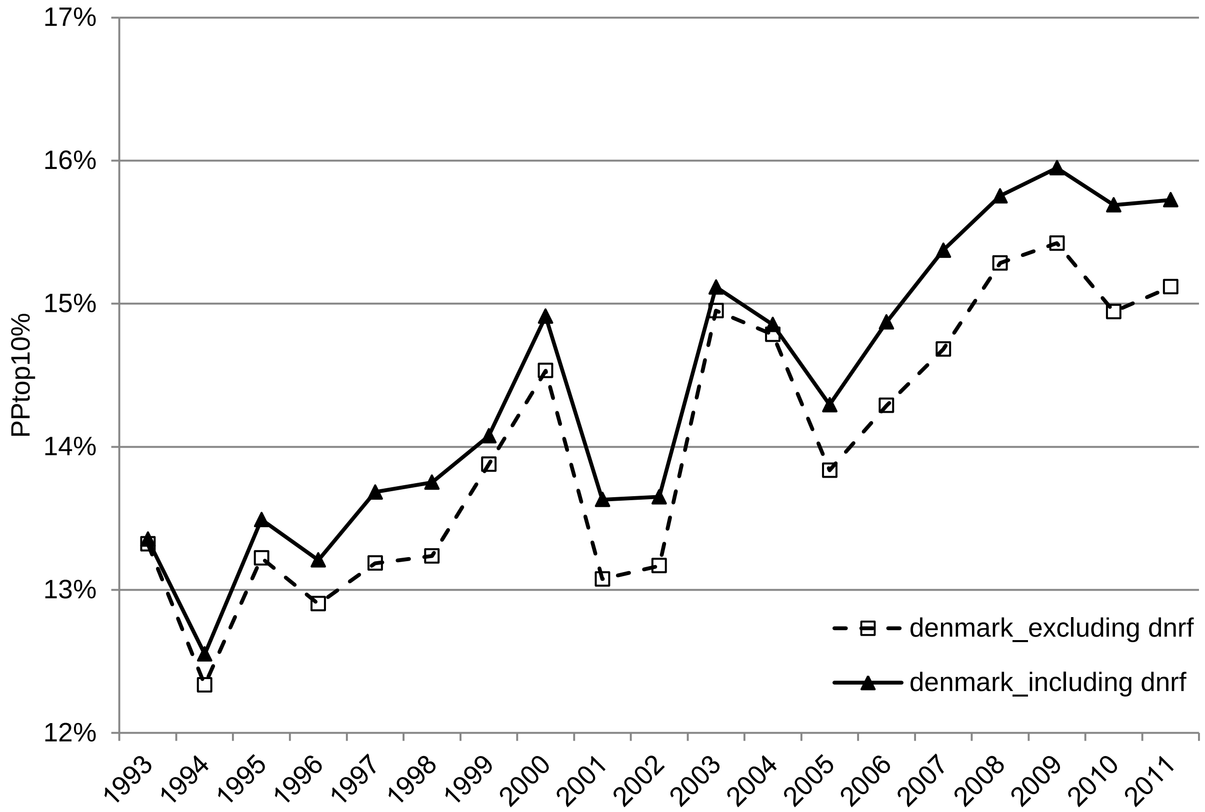



The figure elaborates on Figure 3.2, by showing the actual numbers involved. It is clear that Danish output more than doubles during the period; however one should take into account the general increasing number of journals and their corresponding rise in volume size in WoS when reflecting on these numbers.  We know from other studies that Danish output in this period more or less follows the general development in the database. Interestingly, the development in the share of DNRF-publications after five years reaches 7.6% and then with some annual fluctuations reaches 8-9% and in the two most recent years analysed 11.5% and 10.8%. In 2007 where the PPtop10% indicator for DNRF peaked at 23.6%, DNRF-publications constituted 756 out of 9277, and where 180.5 of these publications constituted 11.7% of the 1361.7 top 10% highly cited publications from Denmark that year.

Figure 3.5 below shows the development in the PPtop10% indicator for Denmark (i.e., a publication is considered Danish if it has at least one Danish address, notice we use full counting in this analysis and not fractional counting), where we calculate the indicator including and excluding the Danish DNRF-publications, in order to see the difference.

As we would expect, DNRF-publications contribute positively to the overall performance of Danish highly cited papers. The proportion of Danish papers among the 10% most highly cited papers in WoS, drops slightly, between .1 and .7 percentage points, when DNRF-publications are excluded.

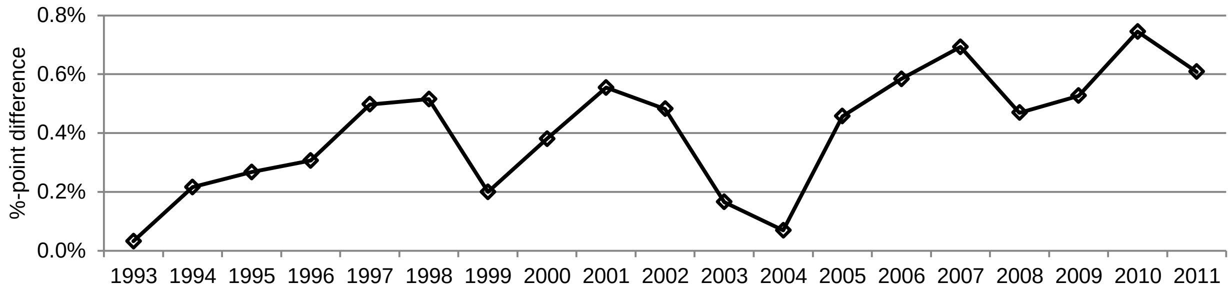
**Figure 3.5: Development in the proportion of Danish papers among the 10% most highly cited papers in WoS, including and excluding DNRF-publications.**




We can also see that as the general share of DNRF-publications rises (see Figure 3.2 and 3.4), where in the latest year's analysed it fluxes around 11% of the Danish output, correspondingly the PPtop10% has risen and from 2007 to 2011 peaks at its hitherto highest numbers. What is noticeable is that the gap between the PPtop10% when including and excluding DNRF-publications seems to be the widest in this specific period. Seemingly, DNRF-publications contribute substantially to the indicator's current level. Figure 3.6 below shows the annual percentage point differences, excluding the inception year 1993; it is noticeable that the low point again is 2003 and 2004.

The average difference for all years, whether including or excluding 1993, is a .5 percentage point difference. Notice also, there is a general drop in Danish performance in 2001 and 2002, though the difference between including and excluding DNRF-publications seems notable. Notice, for the time being, the question whether the effect sizes of the annual drops in the Danish PPtop10% indicator when removing the DNRF-publications can be considered an important or large drop remains unanswered as we basically do not have any reliable benchmark to compare the result with and statistical significance tests are not an appropriate solution (cf. Schneider, 2013). Hence, the claim of a notable difference is based on our impression and experience with similar data analyses.

**Figure 3.6: The annual percentage point difference in the PPtop10% indicator when excluding DNRF-publications from the overall Danish set of publications.**

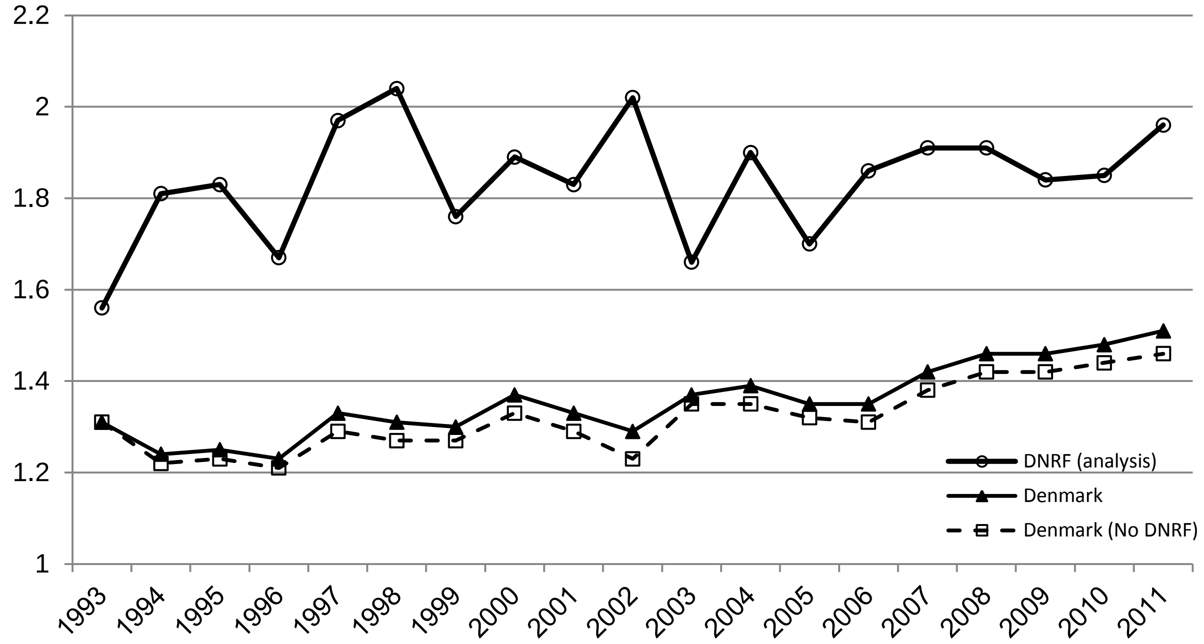



## Mean field-normalized citation scores (MNCS) for Denmark with and without DNRF-publications

As outlined in Chapter 2, the indicators PPtop10% and MNCS are often strongly correlated, especially when publication numbers are high and the statistics are correspondingly robust. However, there can be differences and these appear typically when units of analysis have one or few very highly cited publications that stand out compared to the rest of the publication portfolio. As citation data are very skewed, calculation of average-based indicators are somewhat clumsy, as few very highly cited publications can boost the MNCS indicator, though the interpretation as an average performance of the unit's publication portfolio is problematic. To supplement the overall PPtop10% indicator, we also provide the corresponding development in the overall MNCS indicator for Denmark with and without DNRF-publications in Figure 3.7 below.

As in the previous analysis with the PPtop10% indicator, here we can also see a higher performance in terms of the MNCS indicator for the set of DNRF-publications compared to that of Denmark as a whole. On the other hand, again it is possible to see a slight steady increase in the MNCS value for Denmark, while in the case of the DNRF we find more variation in the value of MNCS over time, although in general terms there is an increasing pattern for the DNRF, with a fluctuation of MNCS values between 1.5 and 2. Also, we can see that DNRF-publications quite consistently contributes with an extra .04 indicator points to the Danish MNCS-indicator

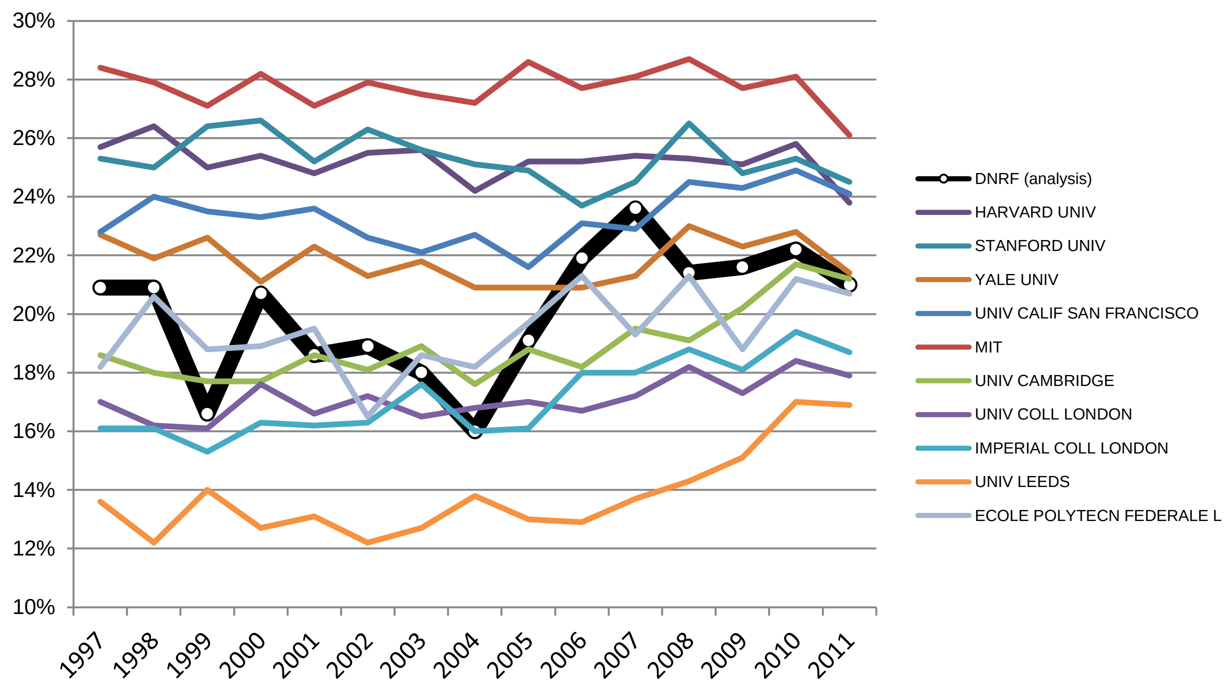
**Figure 3.7: Development of the MNCS indicator for DNRF, as well as Denmark with and without DNRF-publications.**




## Developments in highly cited publications for DNRF compared to benchmark institutions

In this section the trend analysis of the proportion of the top 10% highly cited publications for the DNRF is compared to the 10 benchmark universities. Notice, the publication window is from 1997 to 2011 in analyses where the benchmark universities are used.

**Figure 3.8: Trend analysis of PP top 10% indicators for DNRF production compared to all benchmark units.**



It is clear from Figure 3.8 that all benchmark universities (as well as DNRF) performs above the expected 10% level; hence these units produce more highly cited paper than expected. It is also evident that some units are performing extremely well – almost continuously above 24% - these are not surprisingly: MIT, Harvard University and Stanford University. When comparing the set of DNRF-publications to the other benchmark units, we can see how in the latest years of the analysis, the share of DNRF top 10% publications has reached the same level as that of benchmark universities such as Yale University, University of Cambridge or the Ecole Polytechnique Federale de Lausanne. Though the annual difference between DNRF and high performing units such as MIT, Harvard and Stanford have been reduced since 2004, the gap is still considerable. Around 2007 DNRF-publications performed as well as UC, San Francisco, though contrary to DNRF, UC, San Francisco's performance have stabilized just above 24%.

When analysing the overall performance of these units, it is important to remember that there is a considerable difference in the size of these units, where size is equal to annual publication output.

Table 3.4 shows the annual size-difference in output between DNRF and the benchmark units (see table caption for explanation).

**Table 3.4: Differences in annual publication output between benchmark units and DNRF. The actual publication output for DNRF is shown in the second column. Columns for benchmark universities show how many times more output a benchmark unit has compared to the annual output from DNRF. For example, in 1997, Ecole Polytechnique Federale de Lausanne had 1.8 times more publications than DNRFs 423, which is an annual output of 778. Harvard University had 20 times more publications totalling 8458 in 1997. The benchmark units are ranked according to their total size-difference compared to DNRF for all years. Harvard is largest.**

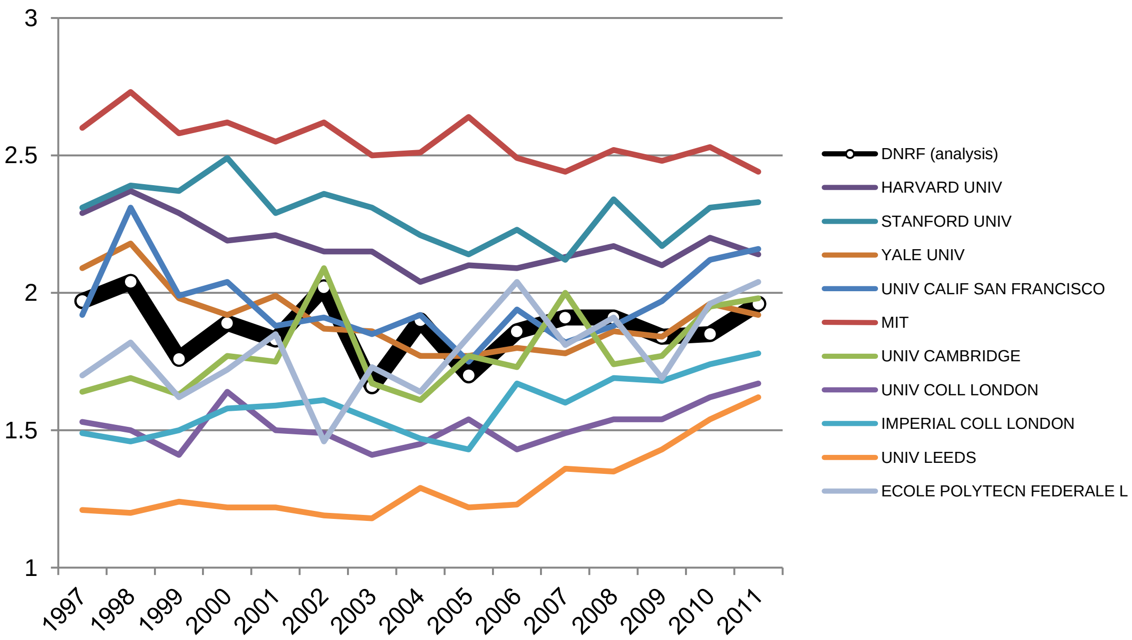
| | No of DNRF pubs | HARVARD | STANFORD | CAMBRIDGE | UNIV COLL LONDON | IMPERIAL COLL LONDON | YALE | UCSA | MIT | LEEDS | ECOLE POLYTECN FEDERALE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1997 | 423 | 20.0 | 8.9 | 9.6 | 7.8 | 6.5 | 7.0 | 7.1 | 7.0 | 3.9 | 1.8 |
| 1998 | 549 | 15.6 | 7.0 | 7.5 | 6.4 | 5.4 | 5.7 | 5.5 | 5.2 | 3.1 | 1.6 |
| 1999 | 554 | 16.1 | 7.2 | 7.8 | 6.4 | 5.6 | 5.6 | 5.3 | 5.5 | 3.2 | 1.7 |
| 2000 | 545 | 16.9 | 7.9 | 8.2 | 6.9 | 6.0 | 5.9 | 5.7 | 5.6 | 3.5 | 1.7 |
| 2001 | 711 | 13.2 | 5.9 | 6.0 | 5.6 | 4.8 | 4.3 | 4.3 | 4.5 | 2.5 | 1.2 |
| 2002 | 678 | 14.1 | 6.4 | 6.4 | 5.9 | 5.4 | 4.8 | 4.5 | 4.9 | 2.8 | 1.5 |
| 2003 | 656 | 14.8 | 6.9 | 7.0 | 6.2 | 5.8 | 5.1 | 5.0 | 5.0 | 3.0 | 1.8 |
| 2004 | 652 | 16.6 | 7.5 | 7.6 | 6.4 | 6.5 | 5.7 | 5.4 | 5.8 | 3.0 | 2.1 |
| 2005 | 759 | 14.7 | 6.7 | 6.5 | 6.0 | 5.6 | 5.1 | 5.1 | 4.9 | 2.8 | 2.0 |
| 2006 | 728 | 16.8 | 7.3 | 7.0 | 6.7 | 6.0 | 5.6 | 5.3 | 5.4 | 2.9 | 2.5 |
| 2007 | 756 | 16.5 | 7.4 | 6.6 | 6.7 | 5.8 | 5.5 | 5.4 | 5.3 | 3.2 | 2.4 |
| 2008 | 789 | 17.1 | 7.4 | 7.0 | 7.0 | 6.1 | 5.6 | 5.3 | 5.2 | 3.2 | 2.5 |
| 2009 | 1013 | 14.0 | 5.9 | 5.8 | 5.8 | 5.0 | 4.6 | 4.5 | 4.2 | 2.6 | 2.0 |
| 2010 | 1267 | 11.7 | 5.1 | 4.8 | 5.0 | 4.2 | 3.8 | 3.9 | 3.6 | 2.2 | 1.7 |
| 2011 | 1326 | 12.1 | 5.2 | 5.0 | 5.0 | 4.4 | 4.0 | 3.9 | 3.8 | 2.2 | 1.8 |
| Total | 11406 | 14.8 | 6.6 | 6.5 | 6.1 | 5.4 | 5.0 | 4.9 | 4.8 | 2.8 | 1.9 |

It is clear that there are some huge size-differences between the units. Harvard University, for example, has an annual output that is higher than the total Danish national output! There is a trend in the data, especially from 2007 onwards. The general difference between DNRF and the benchmark units becomes slightly smaller and this is mainly an effect of a higher increase in DNRF-output than an increase in output for benchmark units.

### Mean field-normalized citation scores (MNCS) for benchmark institutions compared to DNRF-publications

Similar to above, we also provide an MNCS supplement to the comparison of the development in the overall PPtop10% indicator between benchmark units and DNRF-publications in Figure 3.9 below. When compared to the other benchmark units, again we can see how the MNCS values of the DNRF are in the same range as that of units such as Yale University, University of Cambridge or the Ecole Polytechnique Federale de Lausanne, while MIT, Harvard, University of California, San Francisco and Stanford remain as the strongest benchmarks for the DNRF.

**Figure 3.9: Trend analysis of the MNCS indicator for DNRF-publications compared to all benchmark universities.**




It is also noticeable, both in Figure 3.9 and previously in Figure 3.8 (i.e. the corresponding PPtop10% indicator) that at least from 2004 and onwards DNRF-performance seems to be in the intersection between the American and the European universities, where the American universities are generally the highest performing. It should be emphasized here that the journal coverage in WoS is heavily biased towards Anglo-American journals and that this to some extent influences the citation patterns investigated, especially to the advantage of American units of analysis. Nevertheless, units such as Ecole Polytechnique Federale de Lausanne and DNRF are still performing extremely well in this biased Anglo-American publication universe, in recent years at the same level as Yale University.

## DNRF publications' contribution to different percentile classes

In this section we analyse the representation of publication in different percentile classes. These percentile classes are the same as used by the NSF[5] The complete results are presented in Appendix III.

Below, Figures 3.10 and 3.11 show the percentage of publications (accumulative and exclusive) across the percentiles classes 99, 95, 90 and 75.

**Figure 3.10: Percentage of publications (accumulative) by percentile classes (99, 95, 90 and 75) for all units of analysis.**

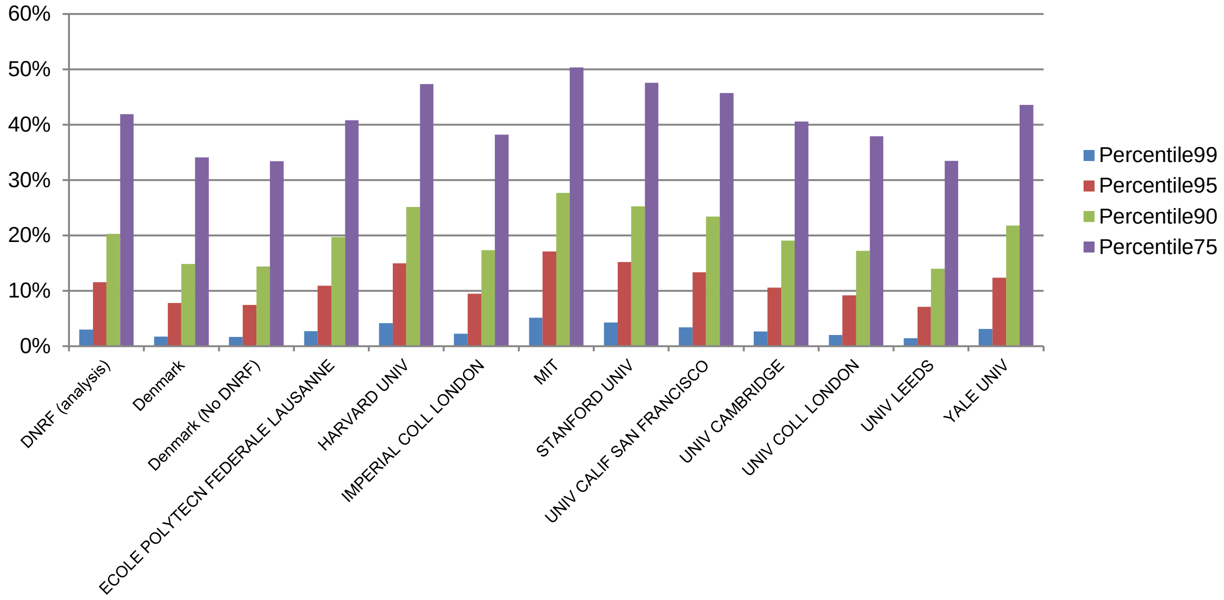


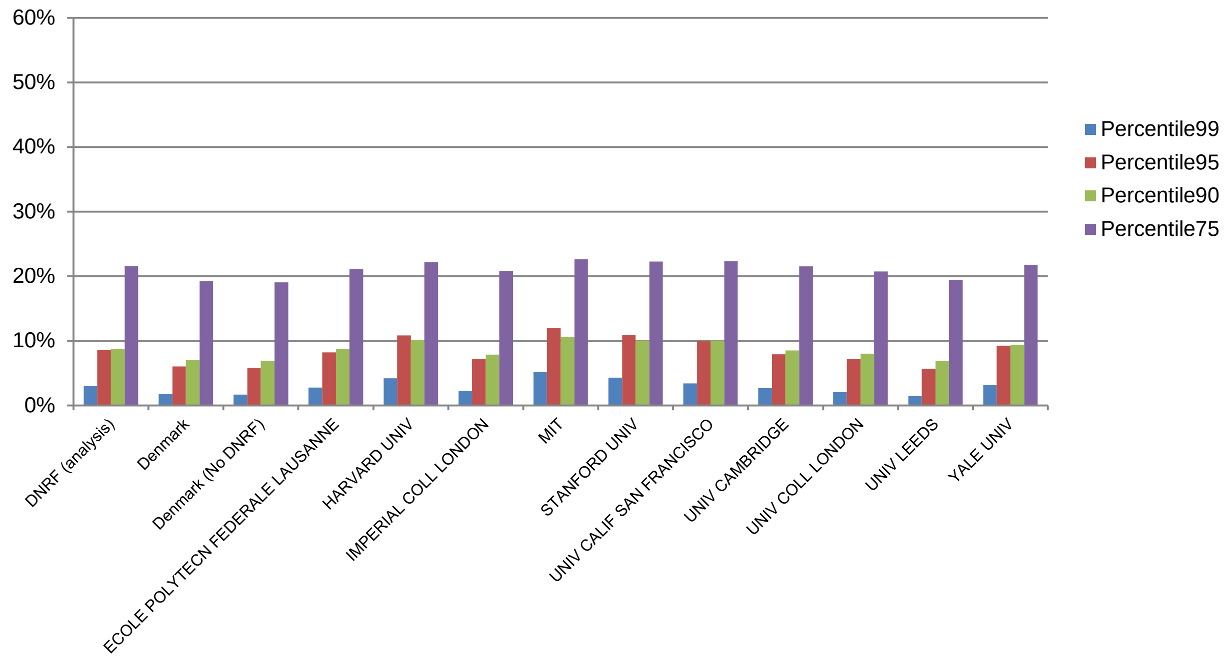

Figure 3.10 shows the performance in terms of the presence of publications across percentile classes. This is the accumulative approach, this means that for every class all the publications in the upper classes are also counted (e.g. for the Percentile 75 class, we also count the publications in Percentiles 99, 95 and 90). Based on this, we can see how in general all the units analysed, proportionally have more publications than expected in the percentile classes. For example, all units have more than 1% of their publications in the Percentile 99 class.

Focusing on the DNRF-publications we can see how they outperform the share of publications per percentile classes compared to that of Denmark and Denmark (no DNRF). Again the strongest benchmark of the DNRF is the MIT that is the unit with the highest values per percentile classes.

Figure 3.11 presents the same result but this time focusing on the exclusive values of the number of publications per percentile classes. In this case, publications are only counted in their highest percentile (i.e. they are counted only one time).

[5] National Science Board. (2010). Science and Engineering Indicators. Washington DC: National Science Foundation. http://www.nsf.gov/statistics/seind10/

**Figure 3.11: Percentage of publications (exclusive) by percentile classes (99, 95, 90 and 75) for all units of analysis.**




In this case we can see how for the DNRF set of publications the share in the percentile 95 class is roughly the same as for the percentile 90 class. For some units (e.g. MIT or Harvard University) the share of publications uniquely in the percentile 95 class outperforms that of the publications in the percentile 90 class. All in all, results are quite consistent with those already presented in Figure 3.10 – all units have a considerable overrepresentation (more than expected) of publications among the most highly cited in the citation distribution.

## Development in the proportion of non-cited publications; benchmark institutions

In this section we explore the evolution of the proportion of non-cited publications among the DNRF-publications compared to the benchmark units. Figure 3.12 presents the trend of the overall PP (uncited) indicators for DNRF and Denmark, including and excluding DNRF-publications. As it can be seen the values at the end of the period increase dramatically due to the shorter citation window for the publications in these years. However, if we focus on the period 1993-2008 we can see how less than 5% of the DNRF-publications receive no citations at all, a quite stable pattern all along the period.

**Figure 3.12: Development in the PP (uncited) indicator for DNRF-publications compared Denmark with and without DNRF-publications.**

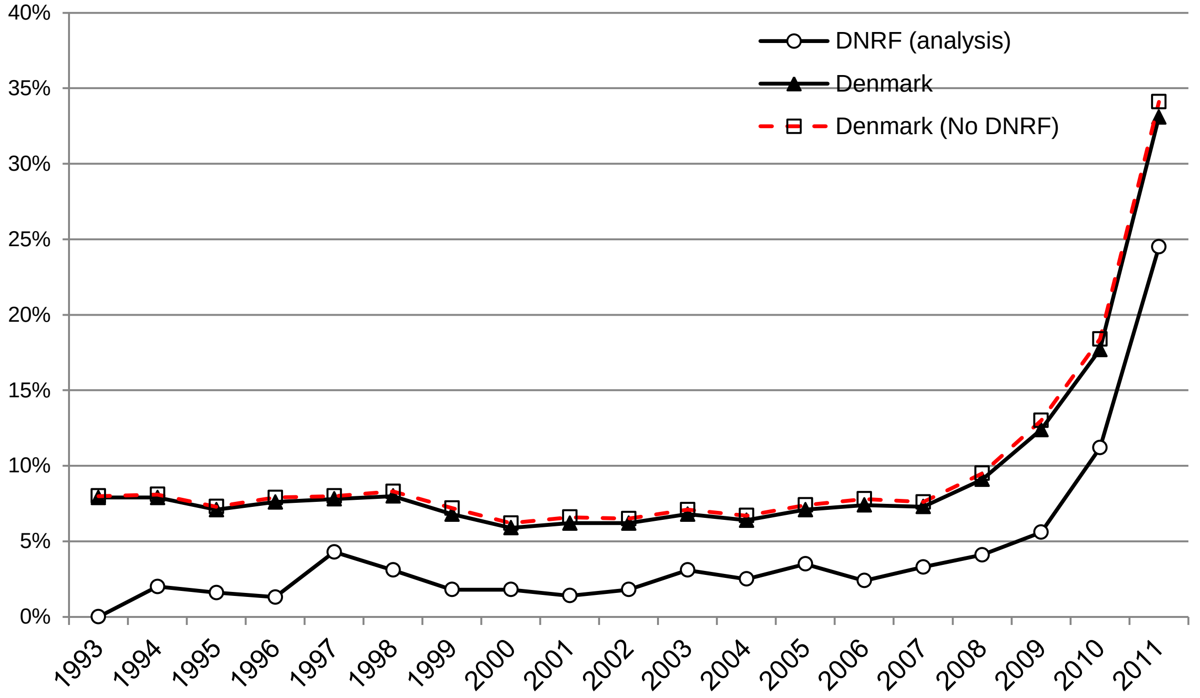
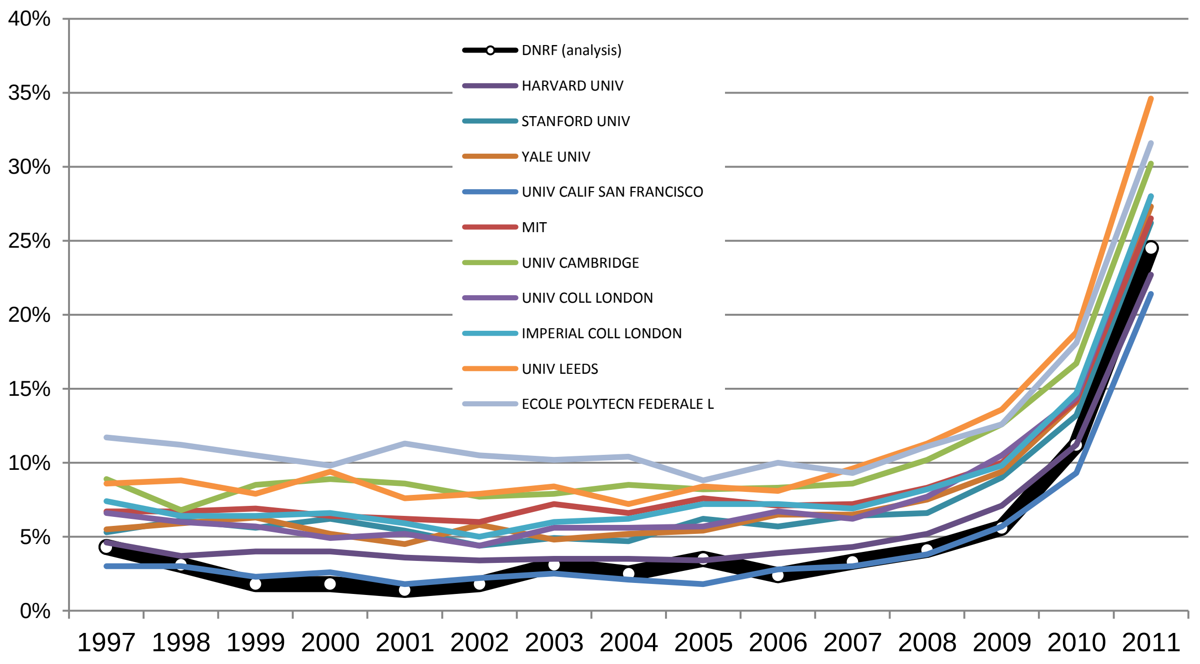



**Figure 3.13: Development of the PP (uncited) indicator for DNRF-publications compared to the benchmark universities.**




When comparing DNRF with the other benchmark units (Figure 3.13 above), it is noteworthy that the values are quite stable over time for most of the benchmark units. It is also noteworthy that the DNRF-publications are among the most less non-cited publications compared to the benchmark units. In essence it is possible to see how the share of non-cited publications of the DNRF is among the lowest (similar level as Harvard or University California San Francisco) being always below 5%, while for most of the benchmark units this value ranges between 5% and 10%. Notice, the increase of uncitedness in the latest years is due to the fact that it takes some time for a publication to accrue its first citation.

# 4. PERFORMANCE IN DIFFERENT SUBJECT FIELDS

In this chapter we breakdown the overall performance of the DNRF-publications into main research fields according to the NOWT-high classification scheme: Engineering Sciences, Medical & Life Sciences, Multidisciplinary Journals, Natural Sciences and Social & Behavioural Sciences. Again we first analyse DNRF-publications contribution to the national Danish impact in these research fields and subsequently we compare the performance of the DNRF-publications with the benchmark institutions. The analysis of the contribution of DNRF-publications to the PPtop10% indicator for Denmark in the aforementioned fields is further broken down into four publication periods. Also, in relation to this analysis, the NOWT-high classification scheme was further broken down to 12 subfields in order to get a more precise picture of DNRF-publications' contribution. The comparison with the benchmark universities is restricted to one publication window including all years from 1997 to 2011.

## Contribution to the Danish PPtop10% indicator in different subject fields

Table 4.1 below shows the difference in the Danish PPtop10% for 5 main research fields in four time periods, including and excluding DNRF-publications. In the previous overall analysis annual PPtop10% were presented. When the overall indicator is disaggregated into the fields resulting in smaller publication numbers within each field, it is recommended to apply longer publication windows. Four periods are chosen with 5 year publication windows, except for the last period where the publication window is restricted to four years.

**Table 4.1: PPtop10% for Denmark including and excluding DNRF-publications in five different research fields according to the NOWT-high classification scheme.**

| | 1993-1997 | | 1998-2002 | | 2003-2007 | | 2008-2011 | |
|---|---|---|---|---|---|---|---|---|
| | DEN including dnrf | DEN excluding dnrf | DEN including dnrf | DEN excluding dnrf | DEN including dnrf | DEN excluding dnrf | DEN including dnrf | DEN excluding dnrf |
| Engineering Sciences | 17.4% | 17.4% | 18.4% | 18.1% | 15.6% | 15.6% | 16.0% | 16.1% |
| Medical & Life Sciences | 12.0% | 11.8% | 13.1% | 13.0% | 14.2% | 14.1% | 15.6% | 15.5% |
| Multidisciplinary Journals | 36.4% | 34.2% | 49.0% | 43.1% | 48.2% | 42.8% | 35.9% | 30.2% |
| Natural Sciences | 15.0% | 14.6% | 14.8% | 14.1% | 14.9% | 14.3% | 14.9% | 13.9% |
| Social & Behavioural Sciences | 8.7% | 8.9% | 9.9% | 10.3% | 11.5% | 11.6% | 13.8% | 13.4% |

It is evident from Table 4.1 that the two main fields where some differences can be noticed when DNRF-publications are excluded, are in: Multidisciplinary Journals and Natural Sciences (marked in grey). It is perhaps not surprising that differences can be seen in these two fields, though it should be emphasized that 'multidisciplinary journals' is a category with relatively few publications, both for Denmark overall and DNRF. However, this does not detract from the fact that the DNRF-publications in this category, noticeably publications in journals such as Nature, Science and PNAS, contributes significantly to the overall impact for Denmark. Natural Science together with Medical & Life Sciences is not surprisingly the two categories where most of the DNRF-publications are classified. It is, however, noticeable that in the category of Medical & Life Sciences close to no difference can be seen in the four periods.

As these two categories are large and somewhat unspecific, we have chosen, for this analysis, to further breakdown the NOWT-high classification scheme into 12 subcategories. Table 4.2 below shows the 12 categories, as well as the PPtop10% of Denmark for each category in each time period, the actual number of Danish publications among the 10% most cited and the share of DNRF-publications among these highly cited Danish publications, and finally the percentage point difference when DNRF-publications are excluded from the calculation of the PPtop10% indicator for Denmark. Notice, a negative percentage point difference means that the Danish indicator drops by that number when DNRF-publications are not included. In each period, the subfields are ranked according to the overall PPtop10%.

**Table 4.2: PPtop10% for Denmark in 12 different research fields in 4 time periods and the DNRF-publications' relative contribution to this national indicator**

| | PPtop10 % for Denmark | No. of Danish top 10% pubs | DNRF's share of Danish top 10% pubs | Percentage point drop in the PPtop10% indicator excluding DNRF pubs |
|---|---|---|---|---|
| **1993-1997** | | | | |
| Multidisciplinary Journals | 36.4% | 95.7 | 12.5% | -2.2 |
| Engineering | 17.6% | 302.7 | 1.7% | 0 |
| Agricultural & Environmental Sciences | 16% | 487.1 | 0.6% | 0 |
| Chemistry | 15.5% | 467.3 | 8.0% | -0.7 |
| Mathematics, Statistics & Computer Science | 15.3% | 224.4 | 1.5% | -0.1 |
| Physics & Materials Science | 15.3% | 751.4 | 10.3% | -0.8 |
| Geosciences | 14.1% | 183.3 | 3.1% | -0.1 |
| Clinical Medicine | 12.3% | 1349.6 | 4.8% | -0.1 |
| Astronomy & Astrophysics | 11.5% | 65.4 | 23.7% | -1.0 |
| Life Sciences & Biology | 10.5% | 771.8 | 3.6% | -0.1 |
| Biomedicine | 10.4% | 540.2 | 6.6% | -0.2 |
| Social & Behavioural Sciences | 8.7% | 87.9 | 88.3% | -0.1 |
| **1998-2002** | | | | |
| Multidisciplinary Journals | 49% | 139.8 | 27.5% | -5.9 |
| Engineering | 18.5% | 403.9 | 3% | 0.1 |
| Chemistry | 16.6% | 616.7 | 13.4% | -1.1 |
| Physics & Materials Science | 15.5% | 871.2 | 10.4% | -0.6 |
| Agricultural & Environmental Sciences | 15.0% | 684.5 | 1.9% | 0 |
| Geosciences | 13.9% | 291.3 | 15.4% | -0.8 |
| Clinical Medicine | 13.9% | 1667 | 7.7% | 0 |
| Mathematics, Statistics & Computer Science | 13.8% | 287.7 | 5% | -0.2 |
| Astronomy & Astrophysics | 12.8% | 104.7 | 25.8% | -0.9 |
| Life Sciences & Biology | 12% | 1105.2 | 7.8% | -0.3 |
| Social & Behavioural Sciences | 10.6% | 160.6 | 56.2% | 0.4 |
| Biomedicine | 10.5% | 637.7 | 11.7% | -0.2 |
| **2003-2007** | | | | |
| Multidisciplinary Journals | 48.2% | 226.6 | 28.7% | -5.4 |

| | | | | |
|---|---|---|---|---|
| Physics & Materials Science | 17.2% | 948 | 12.4% | -0.3 |
| Engineering | 15.9% | 465.3 | 1.3% | 0 |
| Clinical Medicine | 15.6% | 2282.2 | 3.6% | 0.1 |
| Chemistry | 15.5% | 635.9 | 21.2% | -1.3 |
| Geosciences | 14.4% | 343.6 | 12.6% | -0.9 |
| Agricultural & Environmental Sciences | 14.3% | 807.1 | 2% | 0 |
| Astronomy & Astrophysics | 14.1% | 125.2 | 22.4% | -1.0 |
| Life Sciences & Biology | 13.4% | 1358.7 | 15.1% | -0.9 |
| Mathematics, Statistics & Computer Science | 13.3% | 353.4 | 4.8% | -0.1 |
| Social & Behavioural Sciences | 12.1% | 267.4 | 44.1% | 0 |
| Biomedicine | 11.4% | 811.3 | 7.9% | -0.2 |
| **2008-2011** | | | | |
| Multidisciplinary Journals | 35.9% | 282.9 | 37.6% | -5.7 |
| Physics & Materials Science | 17% | 905.8 | 18.9% | -0.8 |
| Clinical Medicine | 16.9% | 2633.9 | 2.2% | 0 |
| Engineering | 16.2% | 508.8 | 0.9% | 0.1 |
| Life Sciences & Biology | 15.3% | 1424.7 | 14.3% | -0.7 |
| Chemistry | 15.3% | 621.4 | 17.4% | -0.3 |
| Agricultural & Environmental Sciences | 14.9% | 834.9 | 4.8% | -0.3 |
| Geosciences | 14.2% | 325 | 19% | -1.5 |
| Astronomy & Astrophysics | 14.1% | 135.9 | 55.5% | -1.7 |
| Social & Behavioural Sciences | 13.7% | 462.9 | 37% | -0.4 |
| Biomedicine | 13.2% | 911.2 | 6.6% | 0 |
| Mathematics, Statistics & Computer Science | 10.6% | 214.4 | 16.5% | -1.0 |

As already indicated by the NOWT-high classification scheme, the contribution to PPtop10% in the Multidisciplinary Journal category is the largest in all four periods. In the last three periods approximately one-third of the highly cited Danish publications in multidisciplinary journals are DNRF-publications. The proportion of highly cited Danish publications in this category is considerably higher than expected and DNRF-publications' contribution to the Danish impact is considerable. More generally we see that in the first period (1993-1997) marked positive contributions come from: astronomy & astrophysics, physics & materials science and chemistry. In the second period chemistry becomes even stronger, physics & materials science perform slightly below the previous period, astronomy & astrophysics is close to stable, but geosciences is new with a strong contribution. In the third period, the performance of physics & materials science almost wanes out, but chemistry, geosciences and astronomy & astrophysics remains strong. New in this period is life sciences & biology with a strong contribution to the national impact for the field. In the final period, physics & material science re-enters with a strong contribution, but chemistry now seems to wane. Noticeable, geosciences and astronomy & astrophysics further strengthen their contribution to the national indicator. Life science & biology is stable, and new is mathematics, statistics & computer science. The generally strong contributions in the last period from most fields should be seen in relation to the general larger number of DNRF-publications in this period –

around 10% of the Danish output analysed in this report. Also, rise and drops should of course be seen in relation to which CoEs are funded and how long their funding runs – a drop could simply indicate that the activity in a field is low due to the termination of one or several CoEs. Finally, the contribution and general impact of these fields should also be seen in relation to the actual publication activity within these fields by DNRF-funded CoEs. The latter is depicted in Figures 5.4 – 5.7 in Chapter 5, where we will discuss the actual activity in relation to impact.

## Profiling DNRF according to main subject fields and comparison of performance in different fields between DNRF and benchmark units

In this section we present the most important result of the performance of the DNRF supported publications compared to all the benchmark units by the main research fields. The table with the main results can be found in Appendix II and here we focus on the most important results and patterns.

In the following figures (Figures 4.1-4.5) we present the comparison between DNRF and the benchmark units for the main research fields in the NOWT-high classification scheme. Each figure presents in parallel both the PPtop10% and the MNCS.

**Figure 4.1: Overall citation impact (PPtop10% and MNCS) of DNRF-publications compared to benchmark units in Engineering Sciences (NOWT-high classification scheme).**

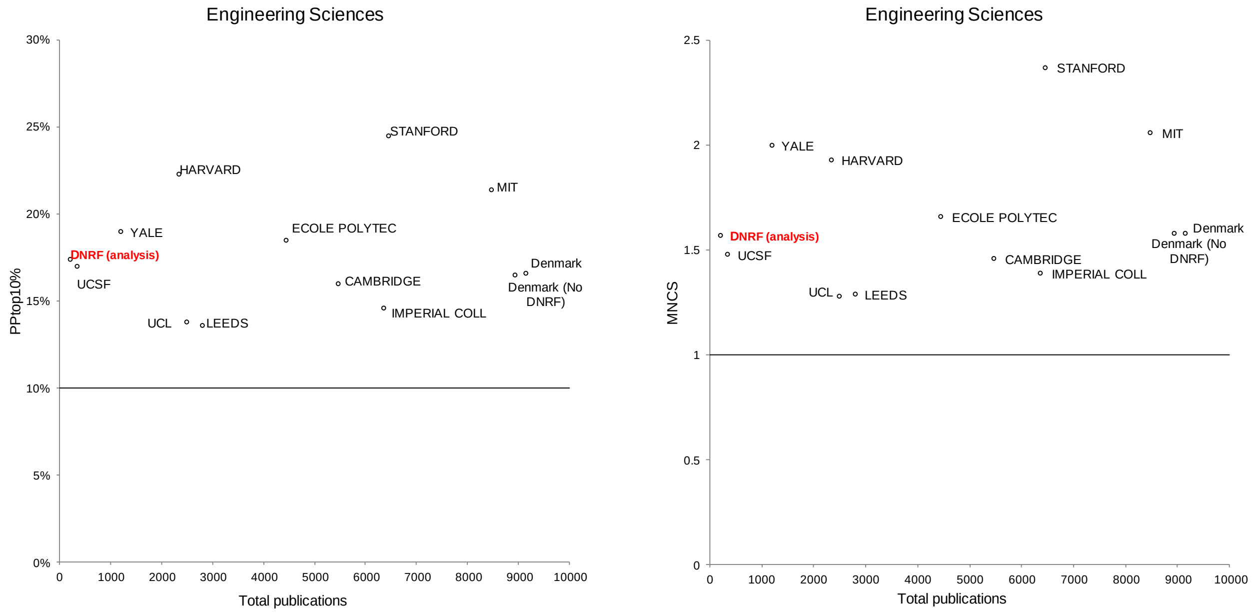



Figure 4.1 presents the performance in the Engineering Sciences where the set of DNRF-publications is the lowest of all the benchmark units, although in terms of impact the DNRF is high and outperforming an important benchmark like the University of California, San Francisco (both in MNCS and PPtop10%).

**Figure 4.2: Overall citation impact (PPtop10% and MNCS) of DNRF-publications compared to benchmark units in Medical & Life Sciences (NOWT-high classification scheme).**

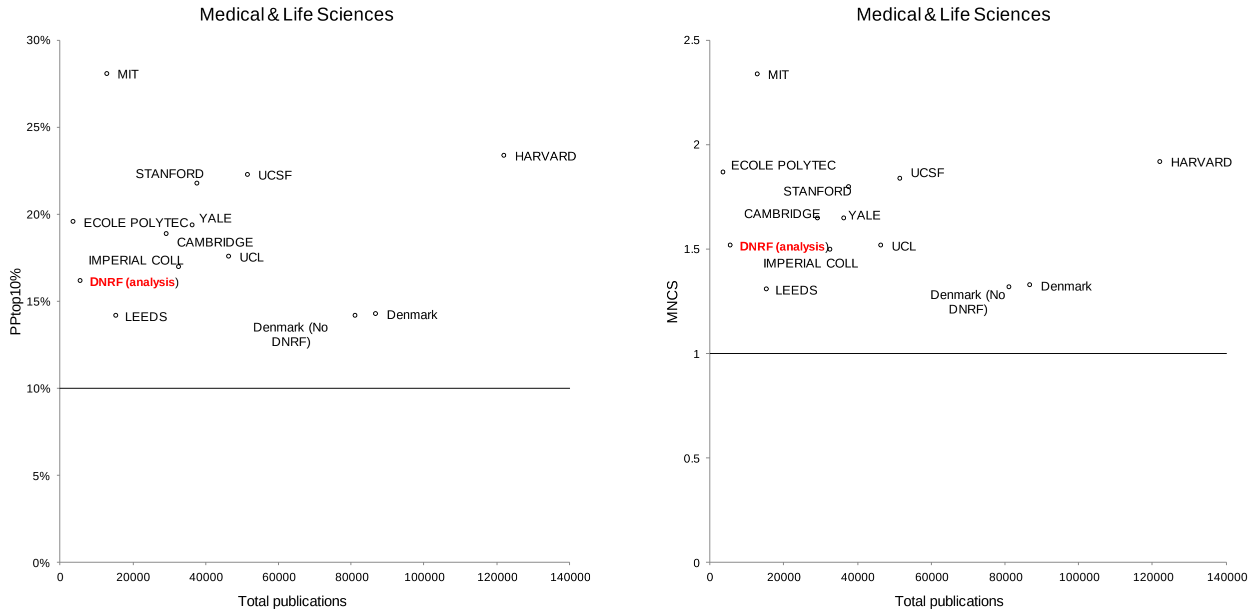



Figure 4.2 presents the results for the Medical & Life Sciences showing that the DNRF outperforms in production the Ecole Polytechnique Federale de Lausanne, while in impact DNRF is at the same level as the Imperial College London and the University College London and outperforming the University of Leeds.

**Figure 4.3: Overall citation impact (PPtop10% and MNCS) of DNRF-publications compared to benchmark units in Multidisciplinary Journals (NOWT-high classification scheme).**

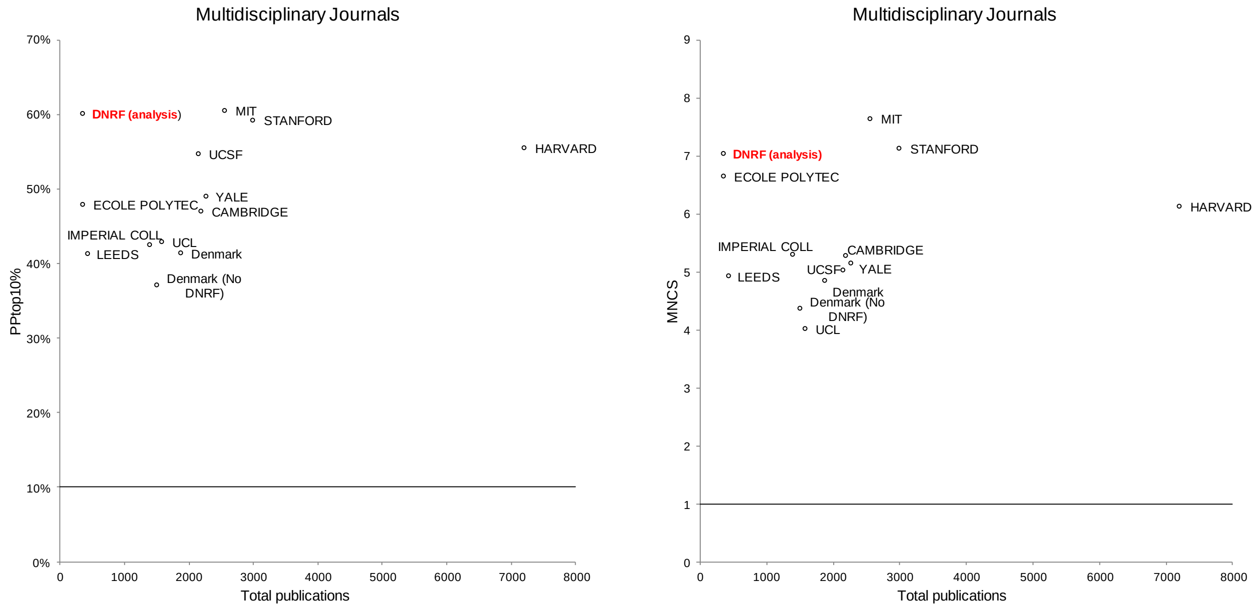


Figure 4.3 shows the results from the ‘Multidisciplinary Journals’ category, while a small category in relation to publication output for the DNRF-set, its impact, both PPtop10% and MNCS is outstanding, performing at the same level as the highest ranked universities in the world!

**Figure 4.4: Overall citation impact (PPtop10% and MNCS) of DNRF-publications compared to benchmark units in Natural Sciences (NOWT-high classification scheme).**

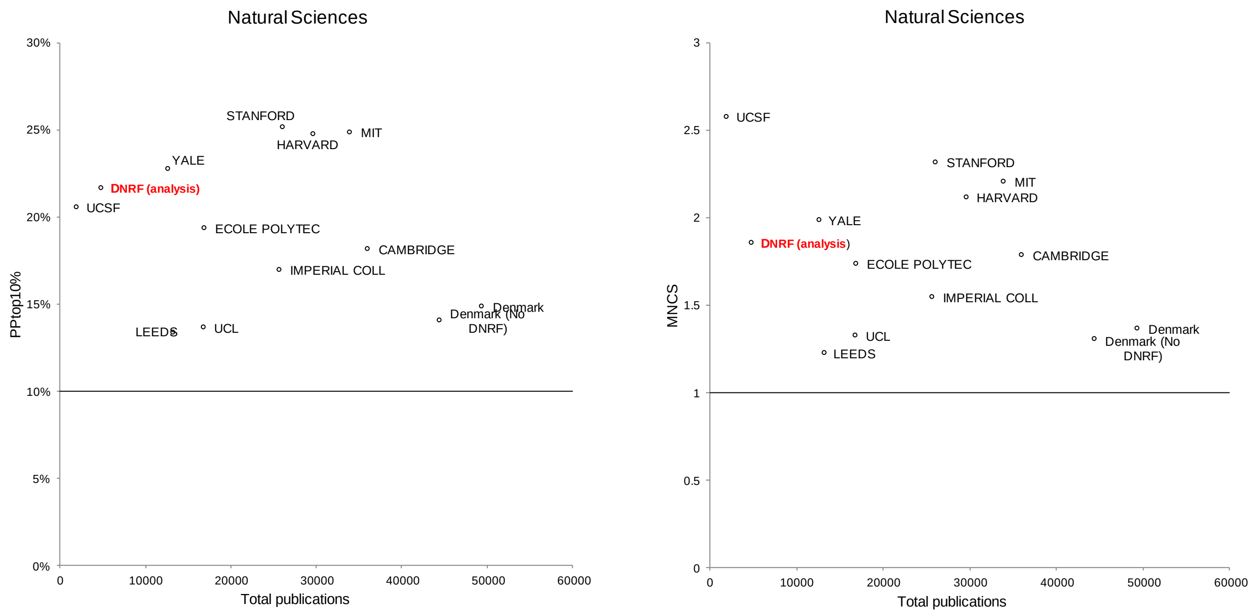
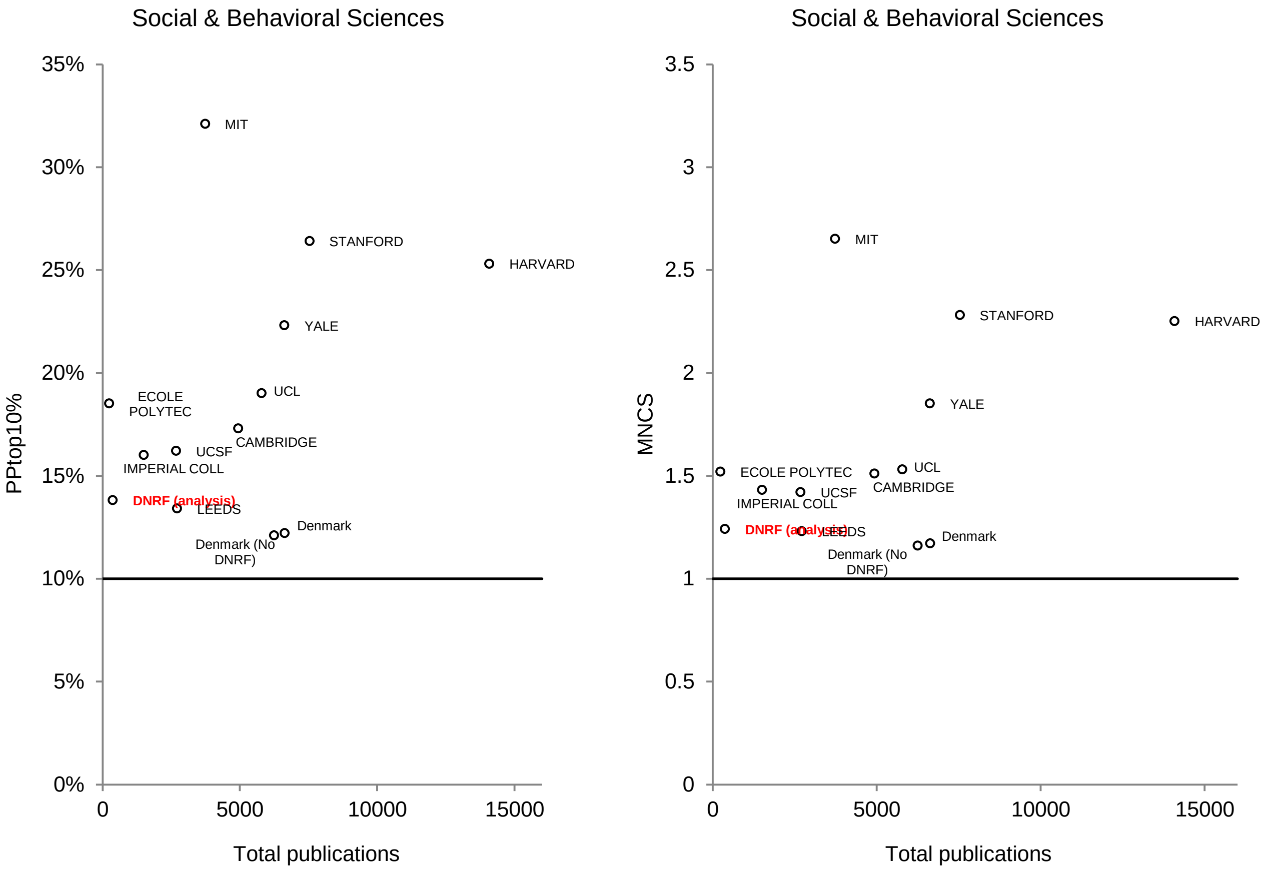



Figure 4.4 presents the results for the Natural Sciences. Looking at the PPtop10% (a more robust indicator regarding outliers) the DNRF is at the same level and even slightly outperforms the University of California, San Francisco, while looking at the MNCS the impact of the University of California, San Francisco is far away from the DNRF. In any case, both in terms of PPtop10% and MNCS the set of DNRF-publications is at the level as that of benchmark units such as Yale University.

Finally, Figure 4.5 presents the results for the Social & Behavioural Sciences, where the DNRF output is comparable to the output of the Ecole Polytechnique Federale de Lausanne although in terms of impact the DNRF is at the same level as Leeds University.

**Figure 4.5: Overall citation impact (PPtop10% and MNCS) of DNRF-publications compared to benchmark units in Social & Behavioural Sciences (NOWT-high classification scheme).**



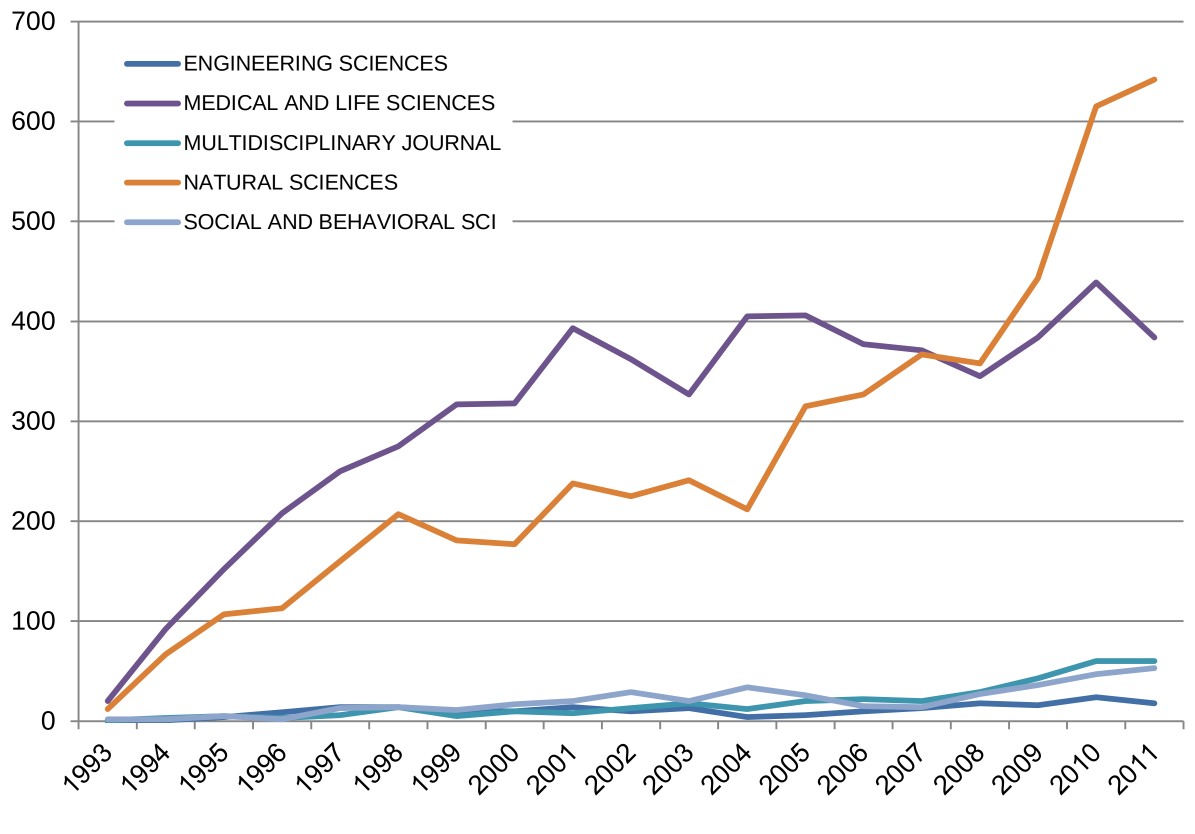
# 5. PUBLICATION PATTERNS

In this chapter we outline the publication patterns for the DNRF WoS journal articles studied in this report. First we outline the development in actual and relative numbers compared to the NOWT-high classification scheme. Subsequently we outline the development in publication patterns in relation to the 12 subfields and the four publication periods used in the previous chapter.

## Development in publication profiles for subject fields for DNRF publications

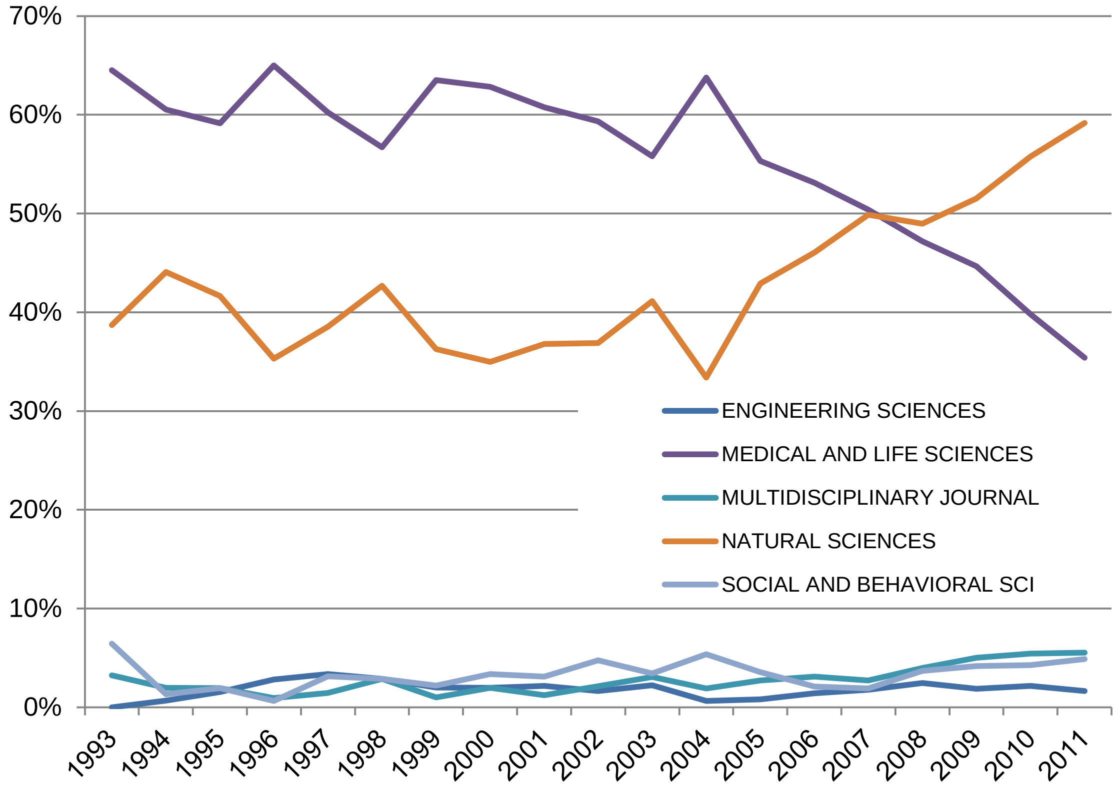
In this section we discuss the development over time of the publications funded by the DNRF for the different fields. Figure 5.1 presents the evolution in the number of publications supported by the DNRF over time for the different fields. As it can be seen, from the beginning the main focus of the DNRF has been on the Medical and Life Science and on the Natural Sciences. In the most recent years, we can notice a shift in the focus, with a stabilization of the production in the Medical and Life Sciences and with the Natural Sciences becoming the most prominent field funded by the DNRF. Figure 5.2 delves into this pattern below.

**Figure 5.1: Annual development in the number of publications per field (NOWT-high classification scheme).**




In Figure 5.2 below we can see the predominance of the Medical and Life Sciences and the Natural Sciences in the profile of the DNRF. In this case it is possible to see how these two main fields actually mirror the evolution of each other, with just a minor contribution of the other fields.

**Figure 5.2: Annual development in the percentage of publications per field (NOWT-high classification scheme).**

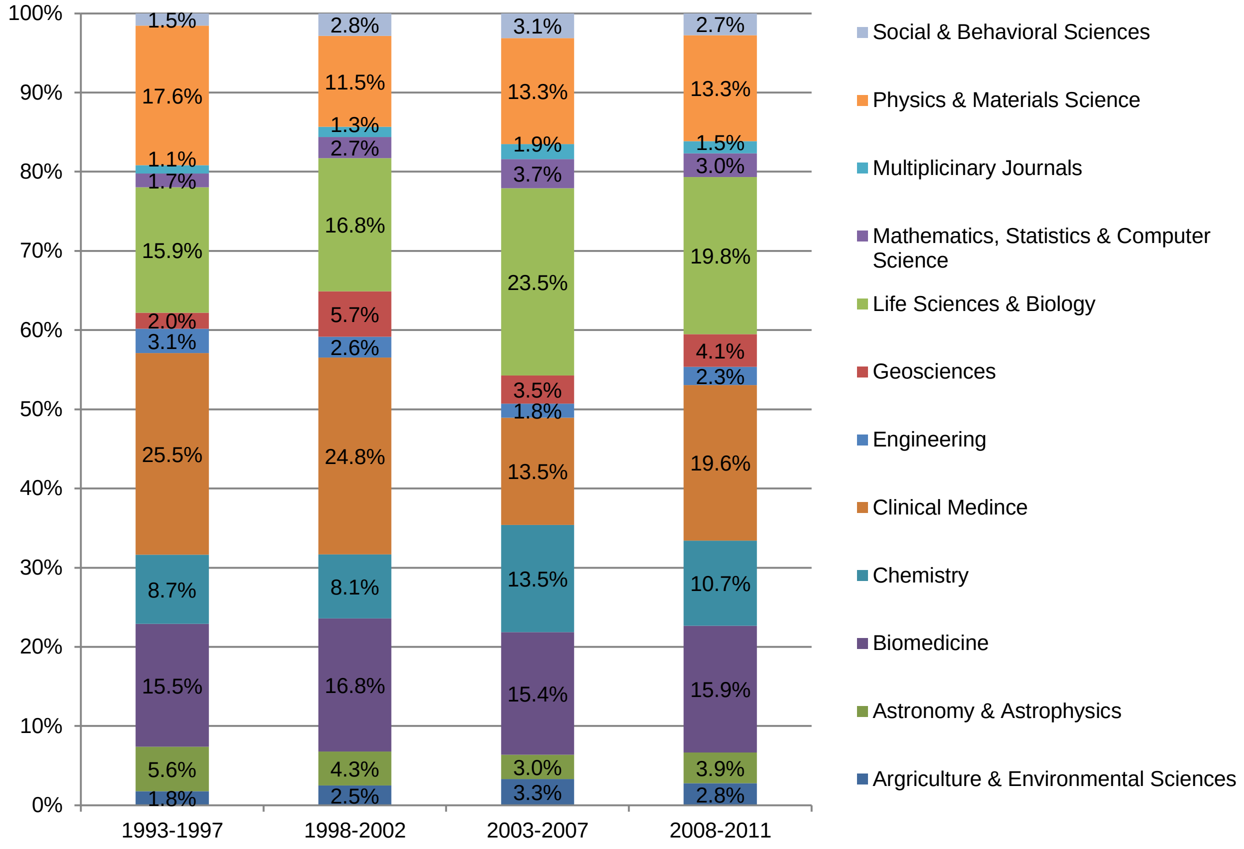



In general the Medical and Life Sciences represents around 60% of the DNRF output while the Natural Sciences accounted for close to 40% of the DNRF supported output. The year 2007 is a turning point in the funding for these two main fields, with both accounting for around 50% of the DNRF funded output, but from then on, the Natural Sciences have the highest predominance in the DNRF output.

Figure 5.3 below outlines the relative publication pattern in four periods mapped onto the 12 subfields previously used in Chapter 4. The largest outputs are seen in subfields such as clinical medicine with 25% of the DNRF-output in the first two periods but then dropping to 13% and in the final period again a slight rise to 19% of the output. Physics & materials science started at 17% but have slightly dropped to around 13% of the output. Life sciences & biology had around 15-16% of the output in the first two periods, but this has risen in the last two periods to approximately 23% and 20% respectively. Publication output in biomedicine has been quite stable around 15-16% of the total DNRF-publication output in the four periods. Chemistry has seen a slight rise from 8% in the first period to 13% in third period and finally almost 11% in last period. Finally, astronomy & astrophysics and geosciences, two fields were DNRF-publications especially in last periods have contributed substantially to the overall Danish PPtop10% indicator, have a relatively small publication activity compared to the other fields where DNRF are active, some 3-5% in the different periods.

Notice that the publication pattern is obviously an effect of the active CoEs at a given point in time and thus eventually of what fields are funded as well as included in this particular analysis.

**Figure 5.3: Relative publication patterns for DNRF-publications in four periods mapped on 12 subfields.**

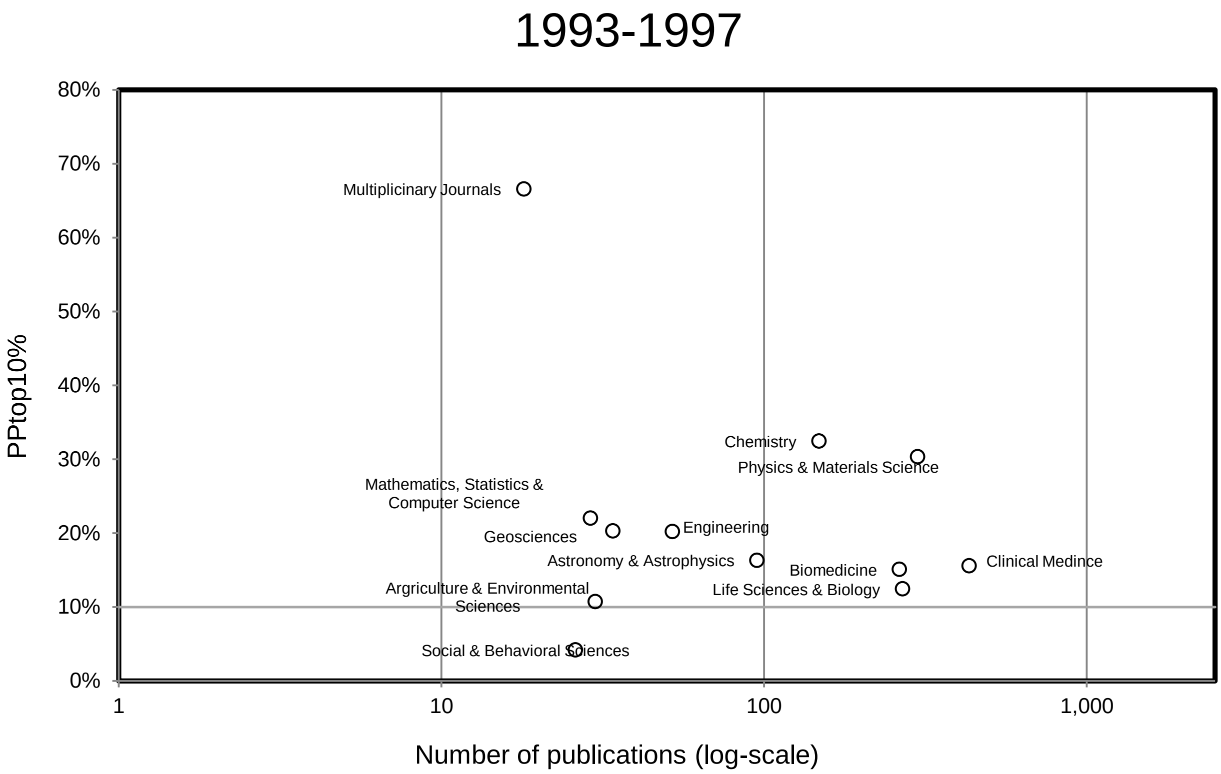



## Publication patterns compared to PPtop10% indicator

To further elucidate the publication patterns in the four periods we have mapped the actual number of DNRF-publications in the 12 subfields as a function of the PPtop10% impact of these subfields, i.e., the impact for the DNRF-publications. The results are presented in Figures 5.4 – 5.7 below. Notice, the x-axis (publications) is log-scaled in order to be able to directly compare the four time periods. As can been seen in the figures, in the latter periods the publication output has generally risen.

As already indicated in the previous chapter, the category ‘multidisciplinary journals’ is by far the highest performing ‘group’ of the 12 subfields. Obviously, the topics of the publications in these multidisciplinary journals actually belong to some of the other categories (and this is factored into the calculation of the indicator values by the CWTS standard overall indicators, i.e. all fields combined). What this high-performing category then tell us is that many more DNRF-publications than expected published in multidisciplinary journals end up being highly cited in their respected fields – in fact around 6-7 times as many as expected. It should be emphasized that while some of the prestigious multidisciplinary journals have much larger citation activity in general, internally they also show a much skewed citation distribution among the individual publications. It also applies for multidisciplinary journals that few publications accrue the majority citations and that many publications receive few or even no citations over the years. Knowing this, it becomes remarkable that DNRF-publications are not only published in these journals, many of them are actually among the prestigious few publications that are very highly cited in these journals.

**Figure 5.4: Number of DNRF-publications from 1993-1997 in 12 subfields mapped as a function of their PPtop10% impact.**

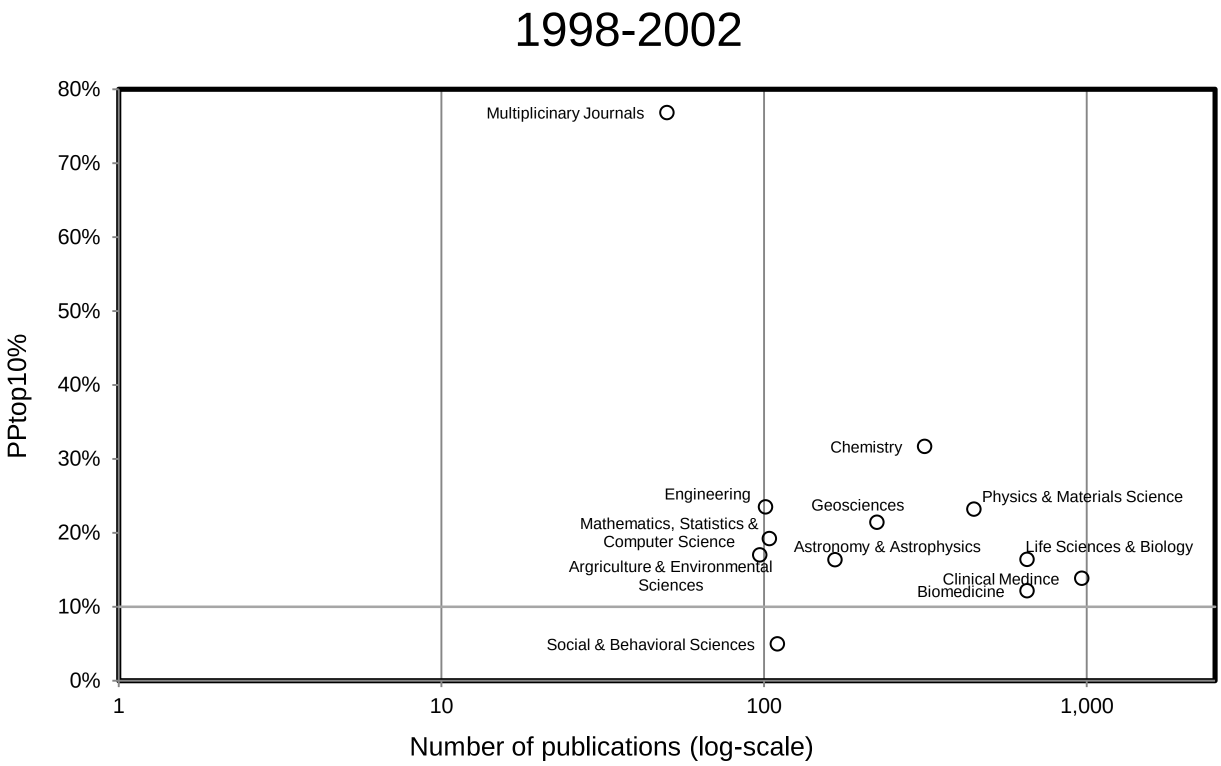



Not surprisingly, in this first publication period where the first funded CoEs were gaining momentum, the publication output was meagre compared to the later periods. This can be seen by focusing on the number of research fields in this period in the 10 to 100 interval on the x-axis and compare this to the later periods in the figures below.

In the previous chapter we also saw that chemistry and physics & materials science were performing very well; in Figure 5.4 we can see the impact mapped in relation to output and here we clearly see that the publication activity in physics & material science is twice as high as in chemistry.

Finally, it is also noteworthy that clinical medicine and to some extent also biomedicine in all four periods are among the most productive research fields, however the general impact of these fields are meagre compared to the other areas (nonetheless above the expected 10% level). This is a general trend that is also visible in scientometric analyses of Danish output. These fields generally account for 1/3 of the annual Danish publication output in WoS, but the impact is not considerably above the world reference value.

**Figure 5.5: Number of DNRF-publications from 1998-2002 in 12 subfields mapped as a function of their PPtop10% impact.**




In this second publication period we see a clear shift to the right in as much as more DNRF-publications are produced in most fields. Obviously, many more CoEs have been funded at this point in time resulting in this publication pattern. Notice the vertical move of geosciences to the right, i.e. more publications but seemingly with similar impact as in the previous period.

In Figure 5.6 below depicting the third period, the output within geosciences remains stable, but now the impact has risen to the second largest in the publication period (after the multidisciplinary journal category).

**Figure 5.6: Number of DNRF-publications from 2003-2007 in 12 subfields mapped as a function of their PPtop10% impact.**

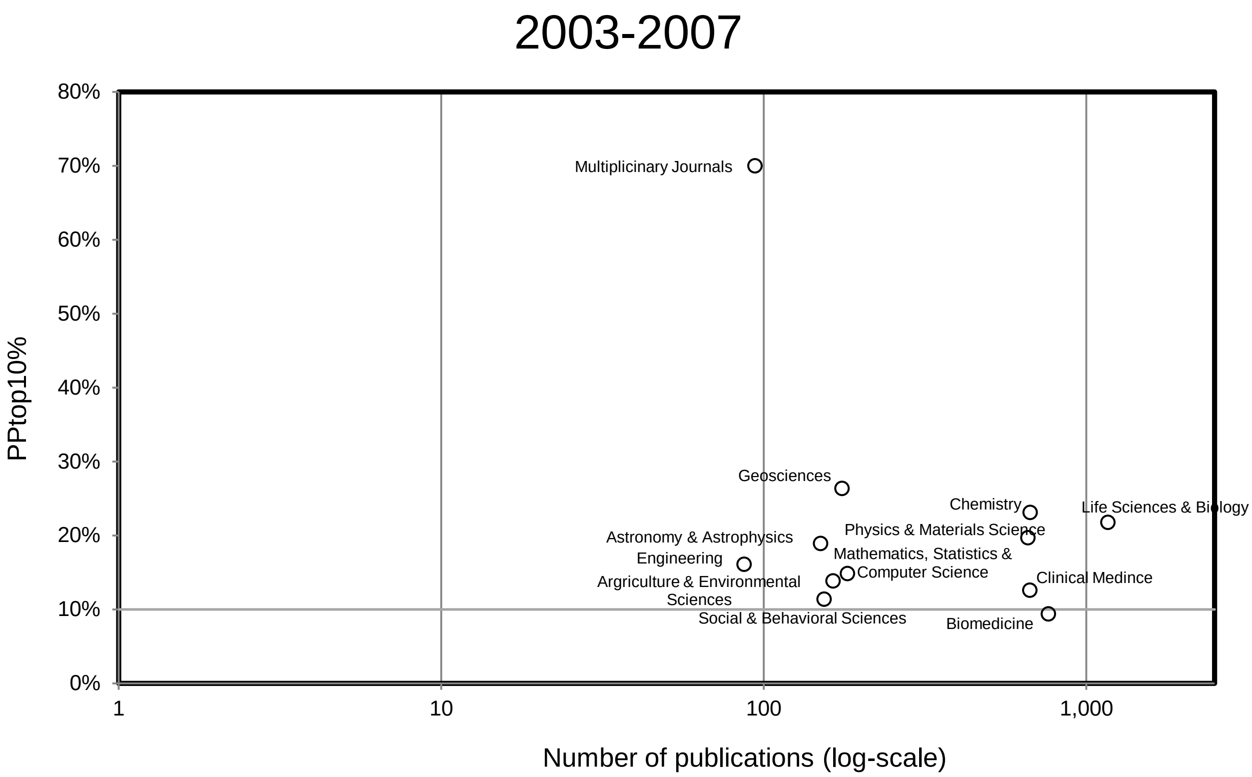
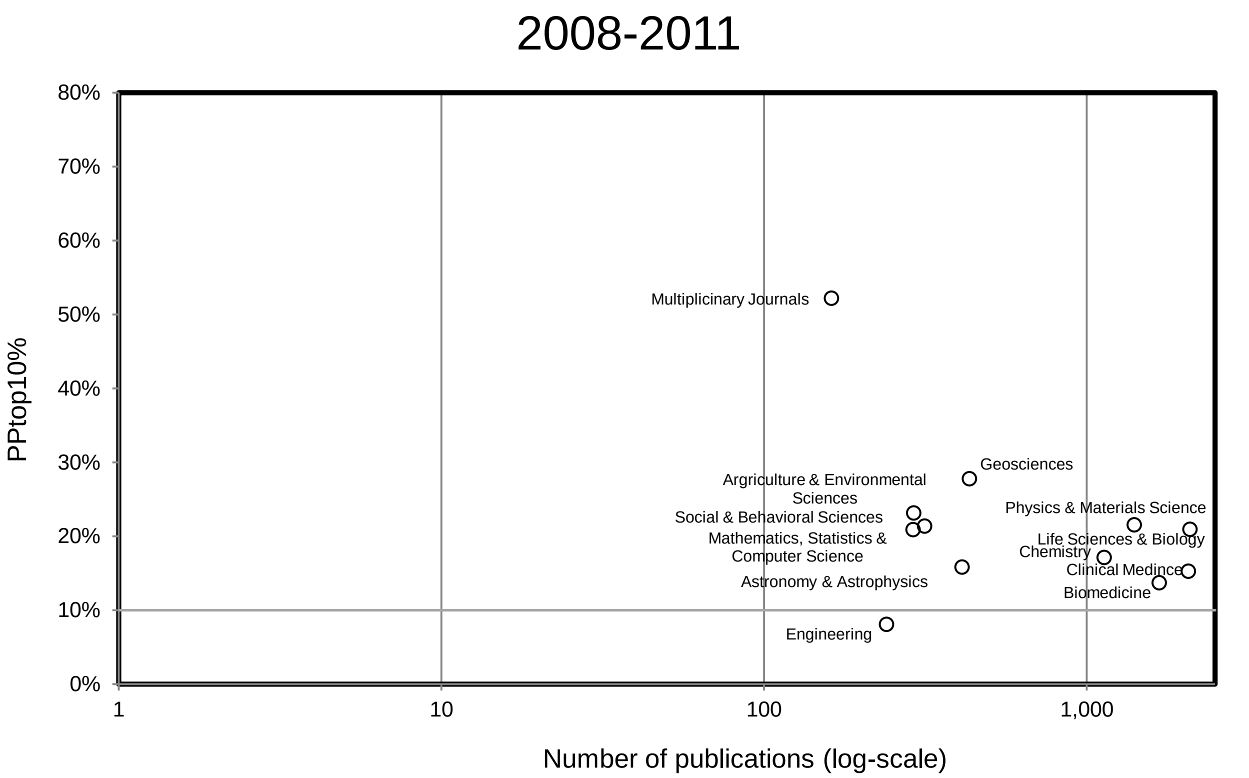



**Figure 5.7: Number of DNRF-publications from 2008-2011 in 12 subfields mapped as a function of their PPtop10% impact.**



A noticeable publication pattern for the last period analysed is that five research fields now have more than 1000 publications and the rest of fields have moved well into the 100 to 1000 interval. Hence in the four periods analysed, publication patterns for all subfields have moved to the right. All fields produce more papers. Notice also that this general pattern is also visible for the multidisciplinary journal category, where in the last period 162 publications belong to this category, however, the impressive impact have fallen slightly in this last period.

# 6. VARIATIONS AMONG CENTRES OF EXCELLENCE

As stated in the introduction the unit of analysis in this report is the DNRF-publications. As such the performance of the individual CoEs included in the analysis is not in focus. However, treating DNRF-publications as a unity, sometimes broken down into fields, can disguise the possible within-difference in publication behaviour and performance between CoEs. It is highly unlikely that 66 CoEs included in the analysis are a homogeneous group. So in order to qualify the previous analyses we have analysed the publication output for each CoE (i.e., the WoS journal publications included in this analysis); the median number of authors per publication for the different CoEs; the citation performance (PPtop10% and MNCS) for the individual CoEs combined for all their publications; comparison of performance between granting periods, and finally two analyses where we first identify the overall proportion of publications on or above the 90$^{th}$ percentile, as well as below this percentile where the Principal Investigator (PI, i.e., leader of the CoE) is author or co-author; and for each individual CoE we analyse the proportion of its publications where the PI is author or co-author, as well as 'leading author' ('leading author' is here defined as being either first or last author; notice in some fields 'leading author', has no implication for the significance of his or her contribution to the publication, for example, in some fields the order of authors in the by line is simply decided alphabetically according to surname).

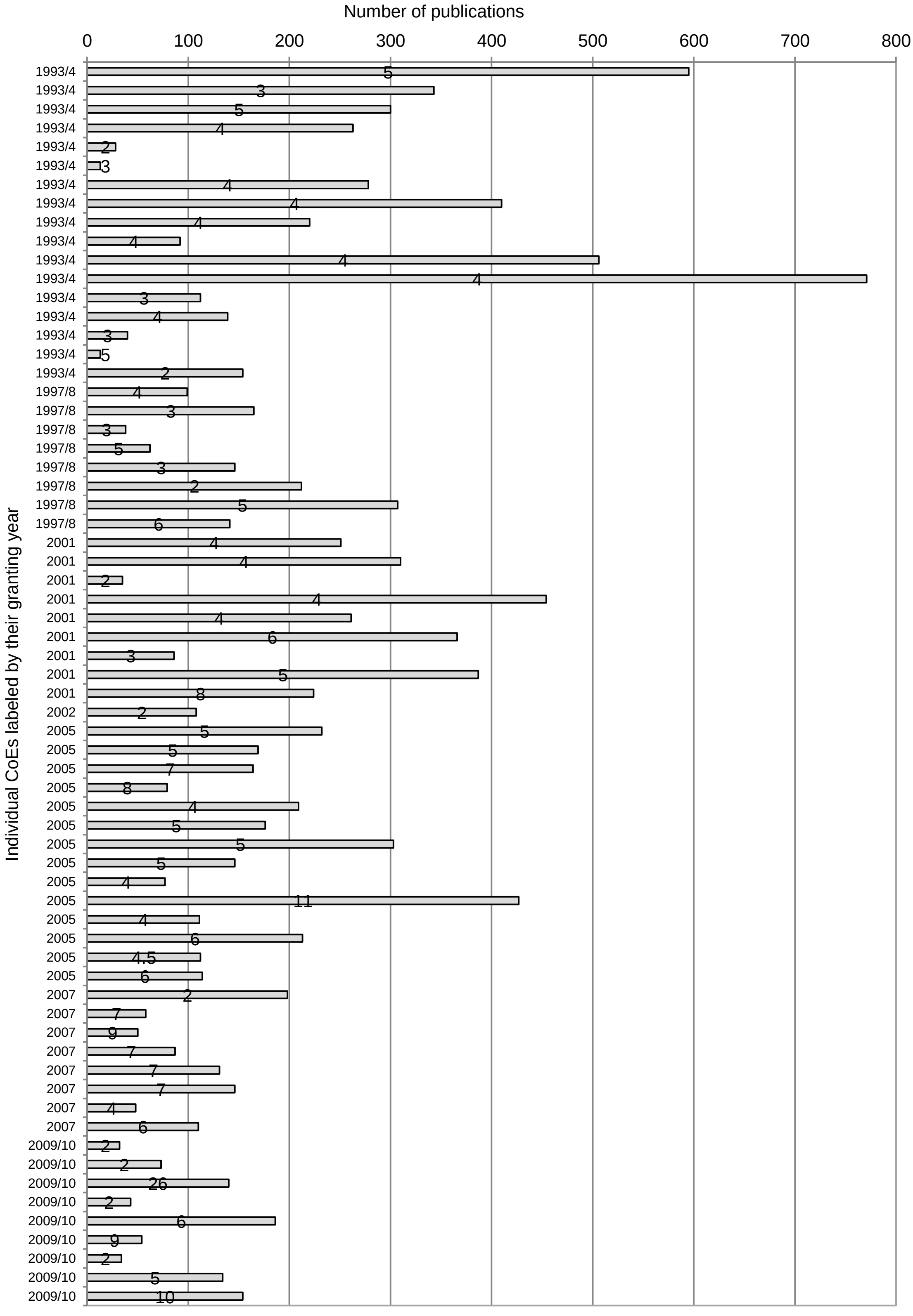
### Publication output for the 66 CoEs and their median number of authors per publication

Figure 6.1 below shows the total number of publications for the 66 CoEs included in the present analysis. The CoEs are group according to their granting year. The number visible in the center of the bars is the median number of authors per publication for this particular CoE. Notice, some of the CoEs in the latter part of the figure is still active and are therefore still producing publications.

Most noticeably in Figure 6.1 is the distinct variation in publication numbers among the CoEs, especially in the upper half and this is interesting because funding to most of these CoEs have terminated. The variation is no doubt due to institutionalized publication behaviors within the research different fields, as well as the coverage of these fields in WoS. Notice, the numbers for the most productive CoE includes publications which are marked as marginally related to the DNRF-funding in the publication list from this particular CoE reported to DNRF.

Looking at the author-profiles for the individual CoEs, we find that the minimum median value among the 66 CoEs is two authors per publication and the maximum is 26 authors. The average median value is five authors per publication per CoE and the median is four. The maximum median value of 26 is an outlier and belongs to a CoE affiliated with experimental physics, a field known to have many authors affiliated to publications. In general, median number of authors is a more robust value than averages as severe outliers can be expected. In the case of the aforementioned CoE with a medium number of authors of 26, the maximum is 3220 authors in one publication and the average number is 716 authors per publication for the particular CoE in question.

**Figure 6.1: Total number of publications for the 66 CoEs included in the analysis. CoEs are grouped according to their granting year. The number in the middle of the bars is the median number of authors per publication for the particular CoE.**



### Variation in citation impact (PPtop10% and MNCS) for individual CoEs

In the present analysis, the funding period for some of CoEs have expired while funding for other CoEs are still running or even in its early phases. Hence, publication output for individual CoEs is not only dependent upon the publication preferences with relation to active field(s), volume and WoS-journal publication, but also their current funding status. In addition, due to the aforementioned irregularities and uncertainties in some publication lists, it is important not to consider the publication output in any absolute sense. It is merely an indication of the publication activity that lies behind the calculation of performance indicators for the individual CoEs. This is important to know, because smaller publication numbers can cause less robust indicators. Hence, in Figures 6.2 and 6.3 we investigate the potential within-difference in indicator values between the CoEs and plot them as a function of publication output. The individual CoEs are marked according to granting year.

**Figure 6.2: PPtop10% as a function of publication output for 66 individual CoEs, where CoEs are marked according to their granting year.**

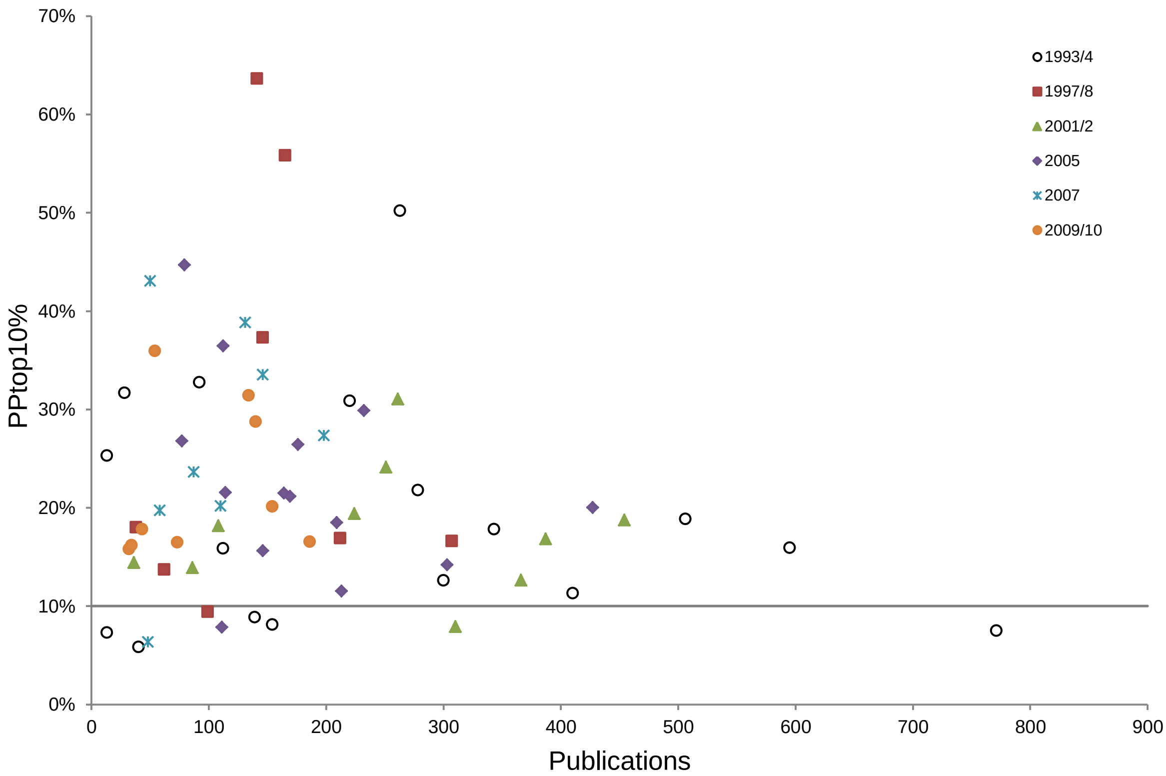
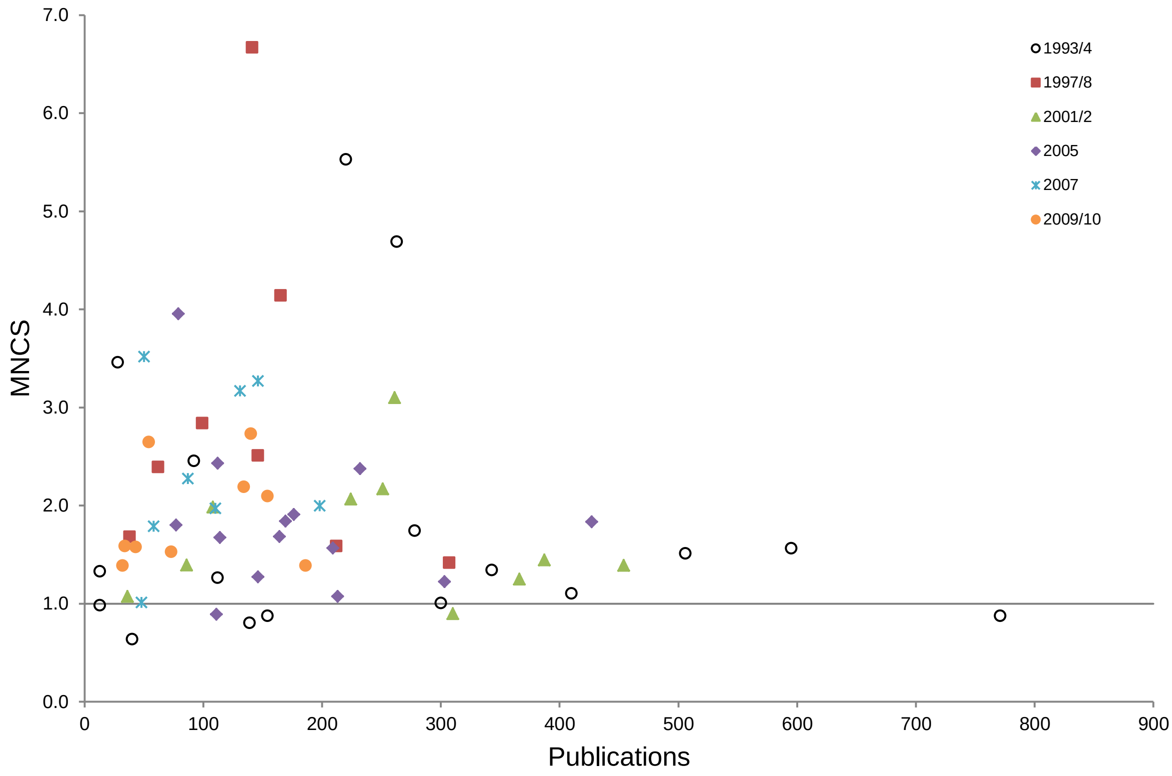


**Figure 6.3: MNCS as a function of publication output for 66 individual CoEs, , where CoEs are marked according to their granting year.**

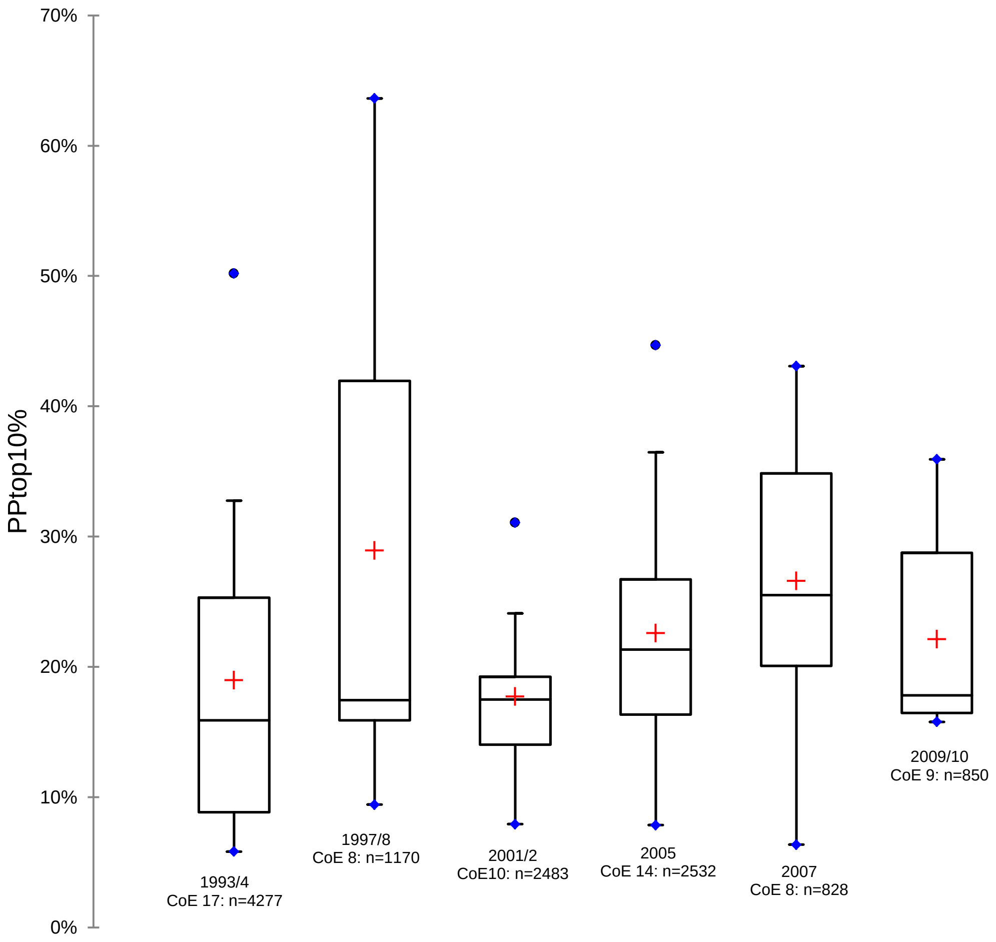



First, it is possible to see how there is no correlation between the PPtop10% or MNCS indicators and the total number of publications. This is not a surprise as it is the general pattern that is commonly found when correlating indicators from different dimensions (e.g. size-dependent [P] and size-independent [MNCS or PPtop10%]).

As expected, considerable variation in output and performance among the individual CoEs can be seen, some CoEs perform extremely well, but others - some 12-13% of the CoEs -perform at or below the international levels for the PPtop10% and the MNCS.

Obviously, some CoEs included in this report are still active and publishing, whereas funding to others has expired. In Figures 6.4 and 6.5 below the CoEs are grouped according to their granting year in order to analyse the potential differences in performance between the different granting periods.

**Figure 6.4: Comparison of the PPtop10% performance between the seven different granting periods. Boxes depict the 1st and 3rd quartiles and the line in the box is the median. The red plus sign is the average and blue dots beyond the whiskers are outliers. Below each box is the number of CoEs in the granting year given as well as the total number of publications from these CoEs.**




Median performance varies to some degree, for PPtop10% (15.9%, 17.4%, 15.5%, 21.3%, 25.5% and 17.8% for each of the seven granting periods), and for the MNCS (1.33, 2.45, 1.42, 1.74, 2.13 and 1.58 respectively). There are noticeable outliers, especially the granting period 1997/8 seems to be affected by this. However, in general it seems that the median level of performance in fact is higher for CoEs funded in the later periods 2005 and 2007, that is CoEs which are most probably still active, compared to CoEs from earlier granting periods where funding has expired.

**Figure 6.5: Comparison of the MNCS performance between the seven different granting periods. Boxes depict the 1st and 3rd quartiles and the line in the box is the median. The red plus sign is the average and blue dots beyond the whiskers are outliers. Below each box is the number of CoEs in the granting year given as well as the total number of publications from these CoEs.**

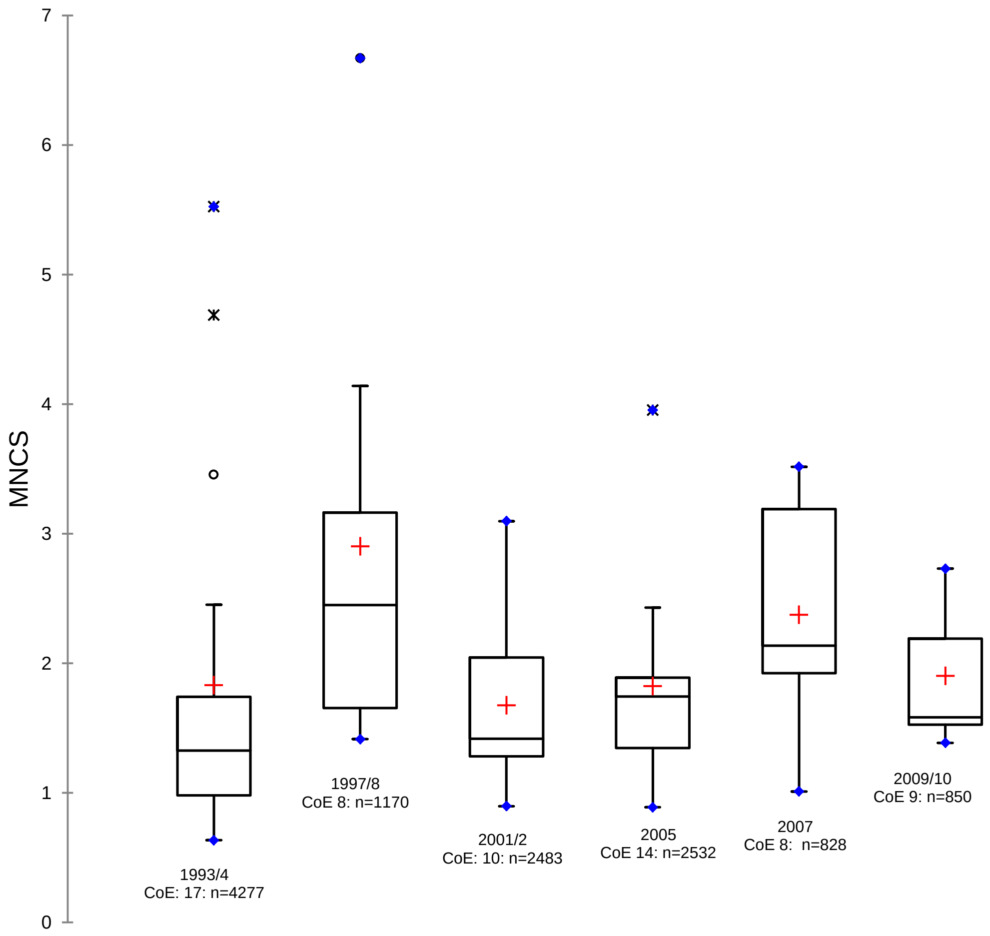


**Figure 6.6: Performance of CoEs where funding expired after 5-years as a function of publication output.**

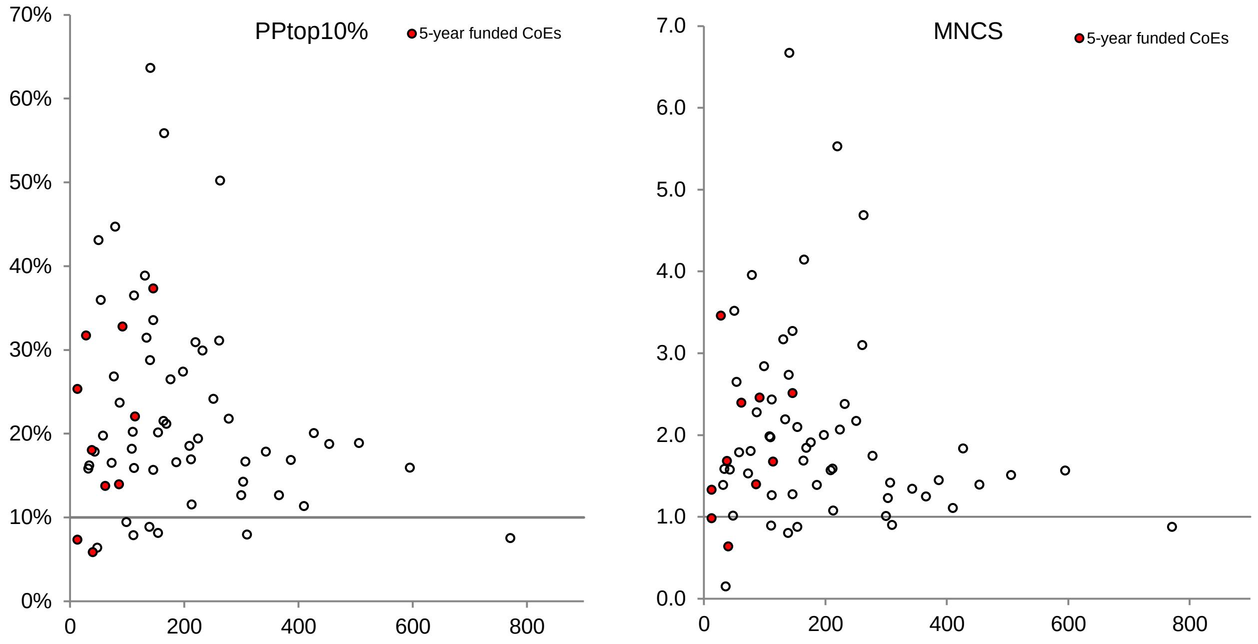



Figure 6.6 shows the performance (PPtop10% and MNCS) for ten CoEs where funding expired after the first 5-year period. Obviously, their publication output is relatively small, whereas their impact varies considerably. Notice, the relatively small number of publications can cause instability in indicator calculations. Various factors can have influenced the relatively small publication output for these CoEs; two of them simply being that the general publication pattern of their field(s) is relatively low and/or that it takes some time from commencing a research project to when the majority of the total number of publications are produced. The latter is visible in Figures 6.7 and 6.8 below, where the cumulative percentage distribution of the publication output for 10-year funded CoEs granted in 1993/4 and 1997/8 respectively, is plotted as a function of time after the initial granting year.

In Figure 6.7 we see three patterns: 1) a few CoEs have produced more than 60% of their publications around year 5 after the granting year (i.e., midterm evaluation); most CoEs have a close to linear publishing rate with approximately 50% of their publications produced around year 5 after the granting year; and finally 3) some CoEs that have a clear time lag in their publication activity, where 50% of their publication output is reached in the second granting period around year 8. This pattern is also visible in Figure 6.8.

**Figure 6.7: Cumulative percentage distribution of publication output for 10-year funded CoEs granted in 1993/4 as a function of time after the initial granting year.**

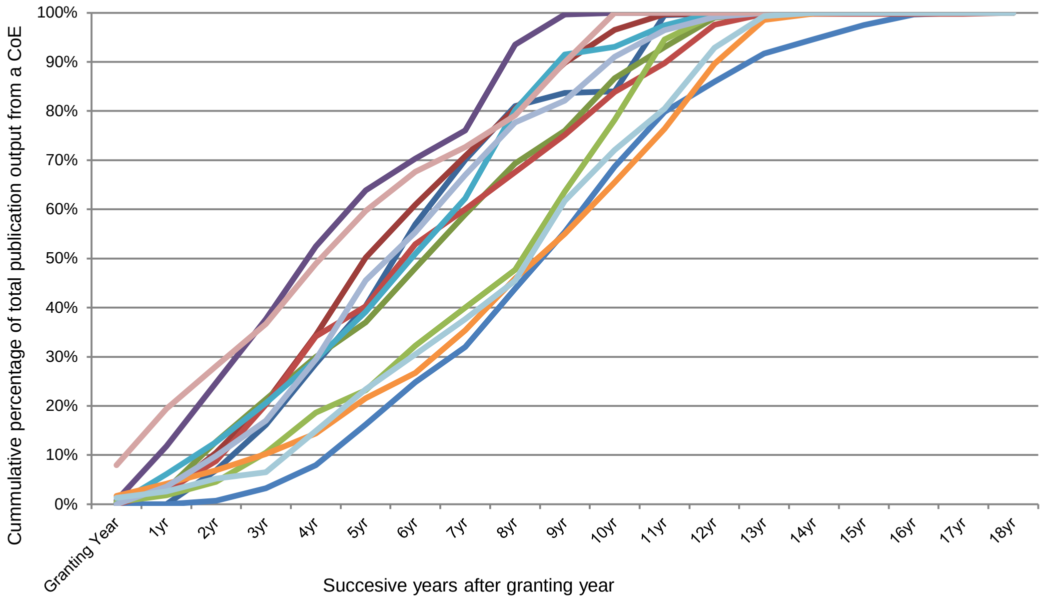
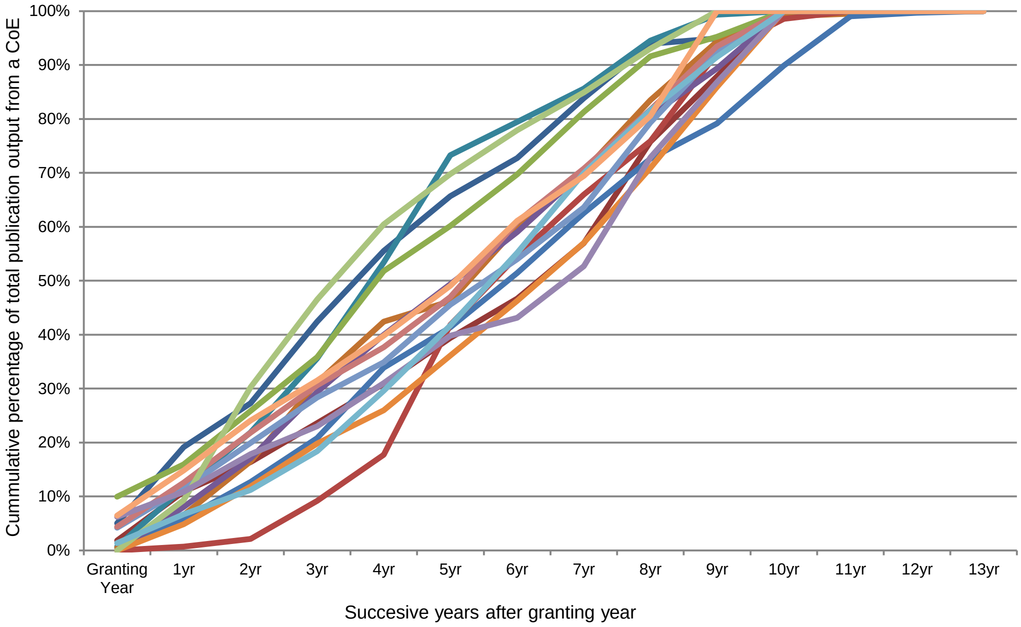



**Figure 6.8: Cumulative percentage distribution of publication output for 10-year funded CoEs granted in 1997/8 as a function of time after the initial granting year.**

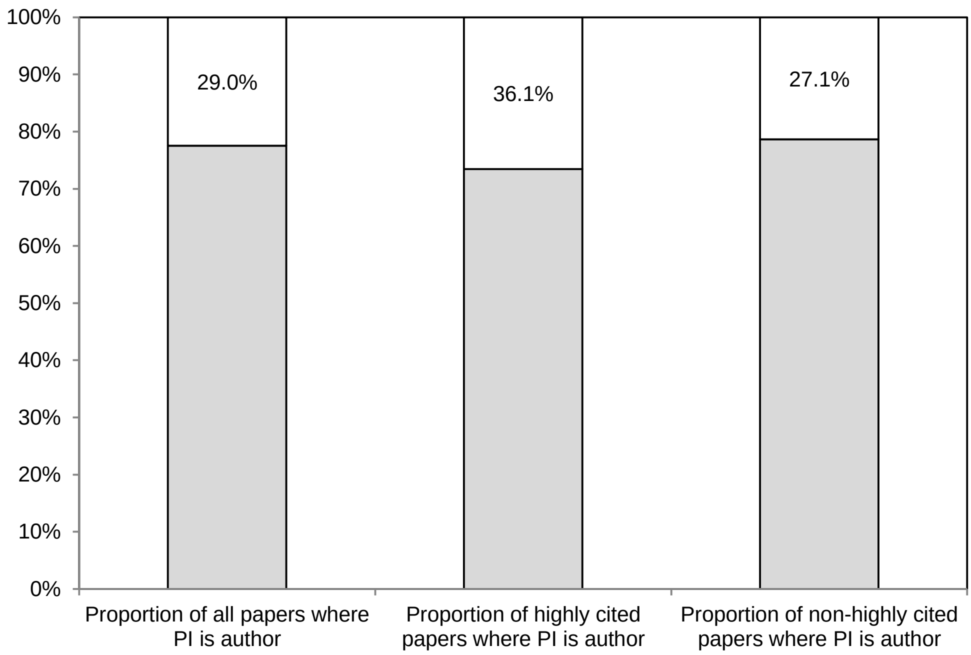


## Proportion of publications and highly cited publications where PI of CoE is author/co-author

In this section we study the proportion of publications and top 10% publications where the Principal Investigator (PI) of the CoE is an author and also when the PI plays a 'leading role' (i.e. appears as first or last author in the author by line of the DNRF-funded publication). In Figure 6.9 we present a summary of the DNRF publications (and top 10% publications) depending on whether there is some contribution of the PI of the CoE.

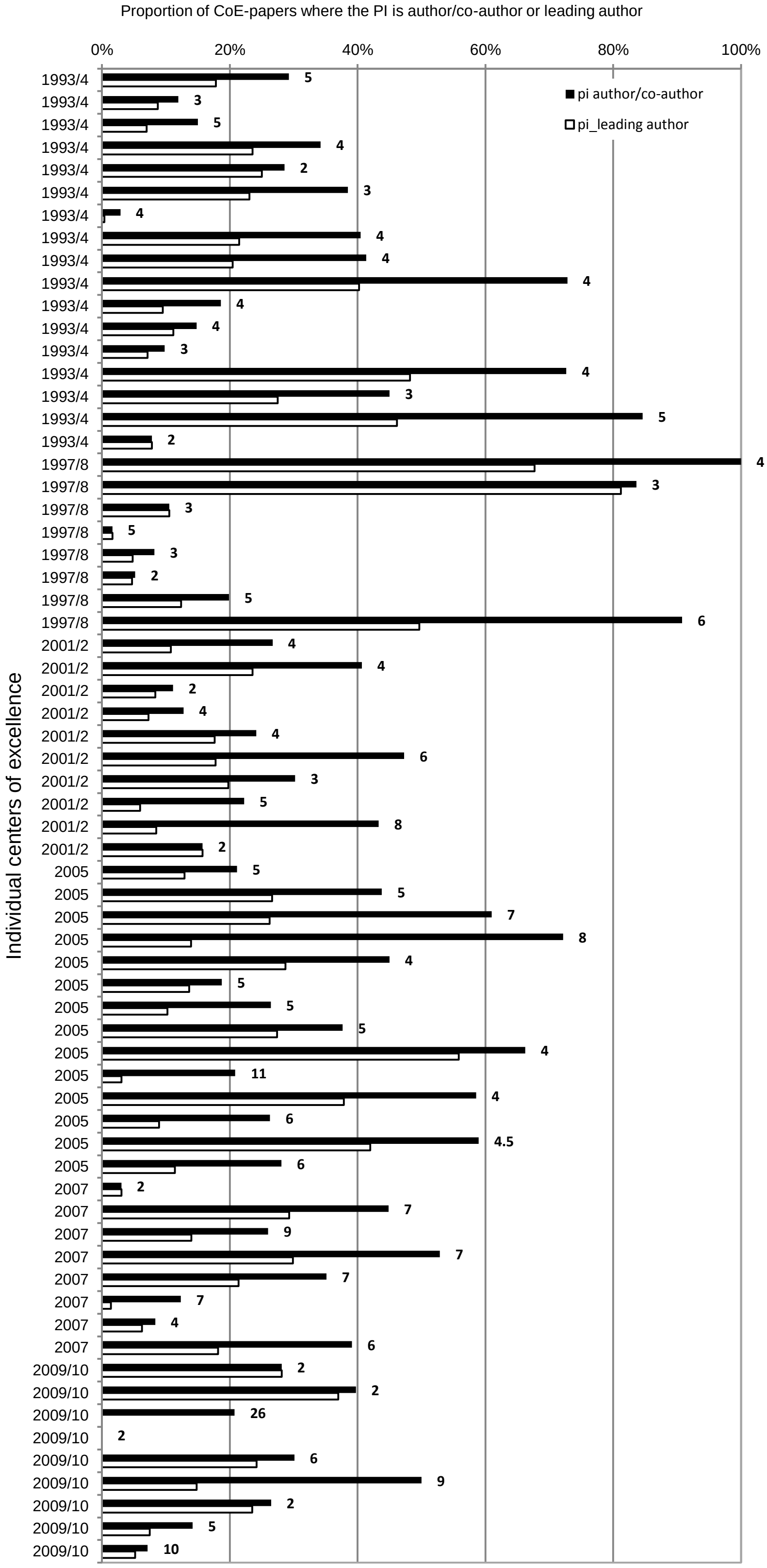
As it can be seen, 29% of the DNRF publications have the PI of the CoE as a co-author, while the PIs participate in more than 36% of the highly cited publications of the CoEs (and around 27% of the non-highly cited publications). In other words, we can state that the PIs of the CoEs contribute slightly more on the highly cited publications of their CoEs.

**Figure 6.9: Proportion of publications and highly cited publications where PI of CoE is author/co-author**




In Figure 6.10 below we present a more detailed analysis of the contribution of the different PIs to the overall output of their CoEs. We can see an important diversity in the participation of the different PIs. In essence most of the PIs participate in less than 40% of the publications of their centres. However, in some CoEs the contribution of the PI to the total output is above 60% and in three CoEs above 80% and in one with 100% (notice these are also centres with a relatively low level of production (<200 publications).

**Figure 6.10: Proportion of CoE-publications where the PI is (leading) author or co-author. The number at the end of the black bars is the median number of authors per publication for the particular CoE.**



We also see that there is much variation between the CoEs when it comes to the PIs' potential role as 'leading' author. We should emphasize again that for some CoEs, the definition of 'leading' author and the assumed role such authors may have is meaningless, however, the individual variation in the pattern makes it difficult to say something in general about so-called 'honorary' authorships to PIs, authorship patterns where the PI seems to credited more for his or her leading role in the CoE than the actual contribution as author to a particular publication.

# 7. AFFILIATION OF PUBLICATIONS TO FUNDING AGEENCIES – NOT ALWAYS STRAIGHTFORWARD

When can a publication be considered ‘belonging’ to a funding agency such as DNRF or affiliated with one of its grants? Obviously, this is by some means a problematic issue. In the initial data collection and processing stages we experienced some discrepancies in the reporting of publications, both in the publication lists produced by the CoEs and in the WoS data. It raises some general issues about reliability. Four CoEs indicated on their publication lists that some publications were in the ‘periphery’ of their the DNRF funding, for example papers authored by visiting scholars not necessarily directly funded by the DNRF grant or directly affiliated to the CoE. We also discovered that some publications (from 2008 onwards) credited DNRF in their funding acknowledgements, though these publications were not mentioned on the CoEs publication lists and there were no Danish address affiliations. The reverse was also detected. Publications mentioned on CoEs publication lists indicating that they are a product of the DNRF-funding; however, DNRF is not credited in the funding acknowledgement in the actual publications, despite that other funding agencies are credited including at least one other Danish agency.

Publications initially ‘belong’ to authors who are affiliated and employed by research institutions. These are normally the units of analysis in bibliometric analyses. However, researchers and research groups, on top of their base funding, often also receive external funding, such as CoE-funding by DNRF. In order to approximate in what respect a publication can be credited as ‘belonging’ to a funding institution and thus by implication be a product of that particular funding, an exploratory analysis of the funding acknowledgements present in WoS publications from 2008 onwards is carried out for the DNRF-publications. As explained in Costas and van Leeuwen (2012), WoS is collecting funding acknowledgment (FA) data from August 2008 onwards. In Table 7.1 we present some of the results regarding this analysis and focusing on the presence of FA within the set of DNRF and Danish publications. The purpose is restricted to analysing DNRF and other Danish funding agencies. In this analysis we include publications until 2012 as there is no citation analysis involved.

**Table 7.1: Funding acknowledgement in DNRF-publications from WoS in the publication period 2008-2012.**

| Publication years 2008-2012 | No. papers identified in WoS | Percentage of total |
|---|---|---|
| DNRF funding acknowledgement alone | 777 | 29% |
| DNRF funding acknowledgement and other Danish funding agencies | 1253 | 47% |
| Papers affiliated to a CoE, with funding acknowledgements to other Danish funding agencies, but *not* DNRF | 658 | 24% |
| Total | 2688 | |

Results in Table 7.1 indicate that around 29% of all the DNRF publications with a FA have an acknowledgement to the DNRF alone, while 47% provide acknowledgments to the DNRF and some other Danish funding agency(ies). Interestingly 24% of the publications from 2008 to 2012 have acknowledgements to other Danish funding agencies, albeit not the DNRF even though the publications are affiliated to a CoE. This perhaps represents a kind of ‘forgetfulness’ of the authors regarding the DNRF (cf. Costas & Yegros-Yegros, 2013 about the idea of ‘forgetfulness’ in FA analysis).

The results are only suggestive but they stress the challenges of ‘ownerships’ of publications and emphasizes that publications can be a result of many influences and several funding organizations. One successful funding often leads to another; this is the well-known phenomenon of preferential attachment, also popularly known as the ‘Matthew effect’.

# 8. COLLABORATION AND IMPACT

In this chapter we analyse the main patterns of collaboration of the set of DNRF publications and we compare it with the patterns of the benchmark units. In the first place, we analyse the main collaboration patterns among all the units and secondly we focus on the impact of the different types of collaboration. Two kinds of collaboration is analysed: 1) collaboration with at least one other national institution (collab), and 2) collaboration with at least one international institution (int collab).

## Share of publications in collaboration with other national and international institutions

Table 8.1 shows how collaboration is a quite common activity for most of the units analysed, with all of them presenting shares of publications in national collaboration above 60%. DNRF shows the highest share of publications in collaboration (73%) and more than 50% of DNRF-publications are done in the framework of international collaboration, showing a similar level as for the Ecole Polytechnique Federale de Lausanne. In this analysis, American universities tend to exhibit a high to moderate level of institutional collaboration (~65%), but a relatively low level of international collaboration (<35%). One should notice that inter-state collaborations in the US are categorized as 'national' collaboration, while in Europe collaboration among neighbouring countries (e.g. Denmark and Sweden) are categorized as international. In any case, British benchmarks also show a lower level of international collaboration (<45%) while DNRF, Denmark as a whole, and the Ecole Polytechnique Federale de Lausanne show levels of international collaboration higher than 50%.

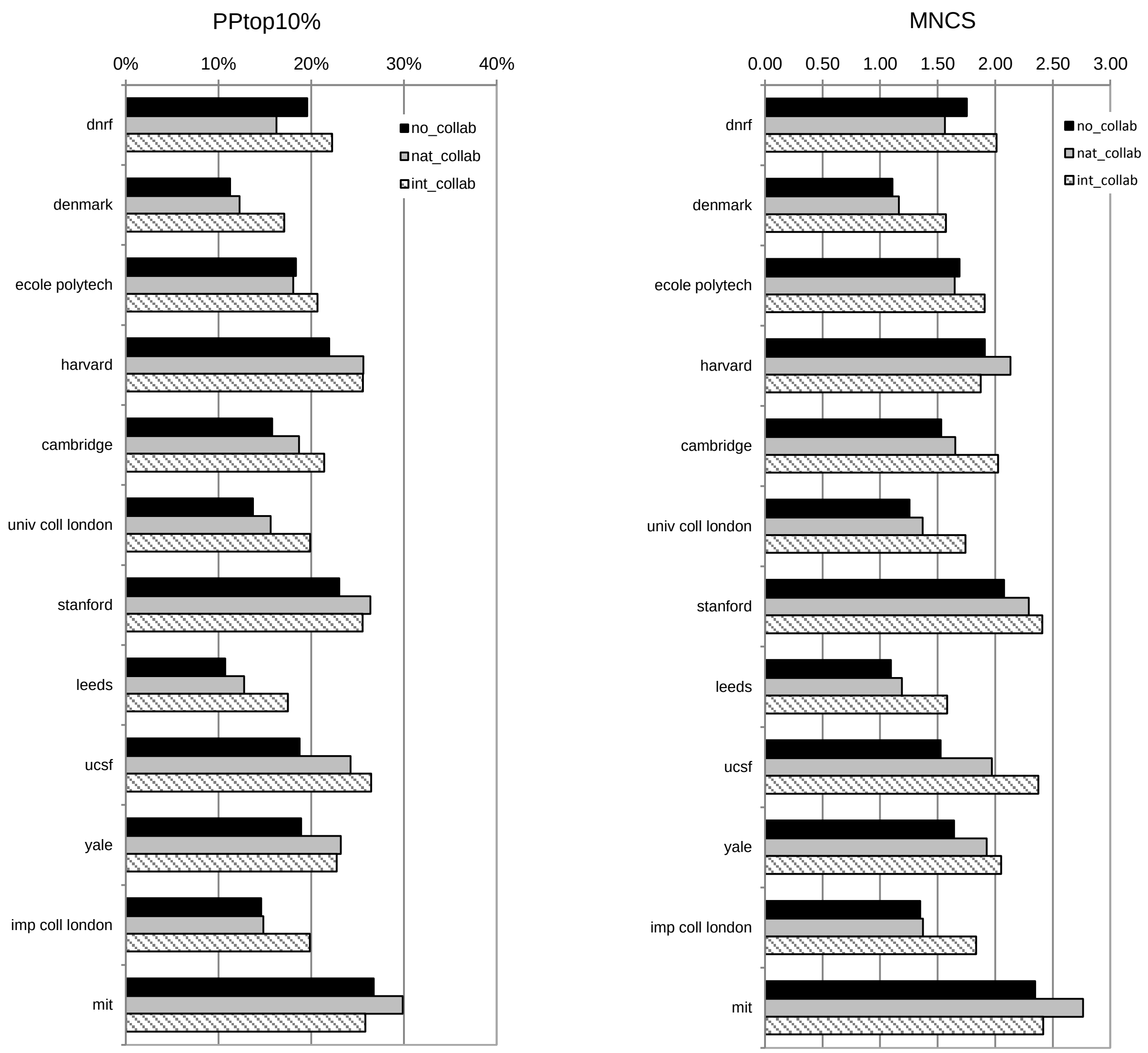
**Table 8.1: Share of publications in collaboration with other national and international institutions; all benchmark units are shown. Units are ranked according PP(int collab), i.e. their share of publications resulting from international collaboration; DNRF are marked in grey.**

| | P | PP(single author) | PP(collab) | PP(int collab) | P(single author) | P(collab) | P(int collab) |
|---|---|---|---|---|---|---|---|
| ECOLE POLYTECN FEDERALE L | 21781 | 4.8% | 69.8% | 57.9% | 1034 | 15202 | 12614 |
| **DNRF analysis** | **10343** | **4.8%** | **73.1%** | **57.5%** | **497** | **7562** | **5948** |
| Denmark | 140894 | 9.5% | 67.2% | 52.4% | 13377 | 94618 | 73780 |
| Denmark (No DNRF) | 130034 | 9.9% | 66.6% | 51.9% | 12854 | 86658 | 67496 |
| IMPERIAL COLL LONDON | 61249 | 6.3% | 70.9% | 47.7% | 3844 | 43425 | 29193 |
| UNIV CAMBRIDGE | 74304 | 14.2% | 63.8% | 45.0% | 10573 | 47414 | 33423 |
| UNIV COLL LONDON | 69257 | 9.2% | 72.3% | 44.4% | 6363 | 50040 | 30745 |
| UNIV LEEDS | 32071 | 11.6% | 62.9% | 36.8% | 3705 | 20162 | 11798 |
| MIT | 55009 | 9.2% | 67.5% | 34.3% | 5050 | 37133 | 18843 |
| HARVARD UNIV | 169131 | 8.9% | 69.3% | 31.6% | 15061 | 117282 | 53409 |
| STANFORD UNIV | 75209 | 10.2% | 66.7% | 28.9% | 7656 | 50183 | 21756 |
| YALE UNIV | 57091 | 12.1% | 65.5% | 28.3% | 6881 | 37390 | 16136 |
| UNIV CALIF SAN FRANCISCO | 55750 | 5.5% | 70.7% | 25.5% | 3056 | 39393 | 14227 |

## Collaboration and impact: PPtop10% and MNCS for national and international co-authored DNRF publications

Publications with international collaboration generally receive more citations than other publications. In this section we analyse the collaboration and impact patterns for all DNRF-publications and compare it with all benchmark units. Figure 8.1 below presents the PPtop10% and MNCS indicators for the two types of collaboration, as well as publication with no collaboration (i.e. only authors from one institution).

**Figure 8.1: Collaboration and impact (PPtop10% and MNCS) for publications with national (nat_collab) or international collaboration (int_collab), or no collaboration (no_collab). The set of DNRF-publications are compared to all benchmark units.**




American universities have a tendency for very strong collaboration with each other (only one in three publications have authors from international institutions) and combined with the general bias in the WoS towards Anglo-American journals, this produces a somewhat different impact pattern than one would expect, where national collaboration often is the highest performing group of publications. The set of DNRF-publications reveal a third pattern. The expected pattern would be relatively highest performance for publications with international collaboration and relatively lowest performance for publications with no collaboration (i.e., authors coming from the same institution). But this is not the case. While publications with international collaboration do have the highest impact around 22% for the PPtop10% for the whole period (some 57.5% of all DNRF-publications have at least one author from an international institution), publications with no collaboration have a PPtop10% value of 19.6% and publications with national collaboration have a value of 16.3%. Somehow, publications from members of CoEs affiliated to the same Danish research institution have considerable international visibility. Among the benchmark universities, only MIT, Harvard and Stanford have PPtop10% indicator values above 20% for the publications with no collaboration.

## 9. JOURNAL PRESTIGE

The previous analyses showed that DNRF-publications belonging to the 'multidisciplinary journal' category performed extremely well. While some journals in this category are obviously considered highly prestigious, a more general analysis of publication and performance in 'high prestige' journals is presented in this chapter. For this analysis, 'high prestige' journals are defined as those journals that have published 30% or more of the top 10% highly cited publications in their fields in a year. In other words, high prestige journals are journals that in a given year-field combination was able to attract and publish almost 1/3 of the 10% most cited publications of their fields. This is a straightforward definition which is free of the limitations of the Journal Impact Factor. A table with main results of this analysis is presented in Appendix IV. It is important to keep in mind that given the fact that we have compared all the units (i.e. DNRF, Denmark and all the benchmark units) this analysis refers to the period 1997-2011. Figures 9.1 and 9.2 present the main development of the number and share of publications in high impact journals (HIJ) over time.

**Figure 9.1: Trend analysis of the number of publications in high impact journals over time for all units.**

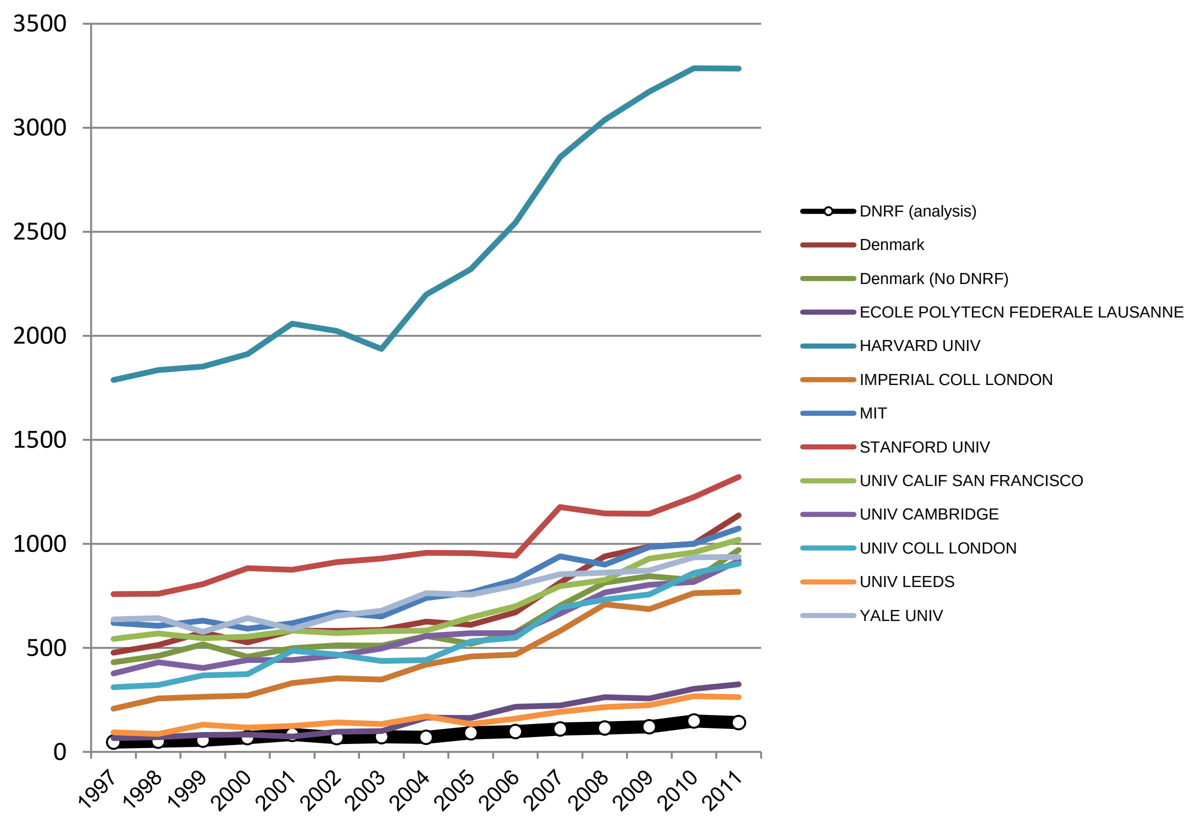
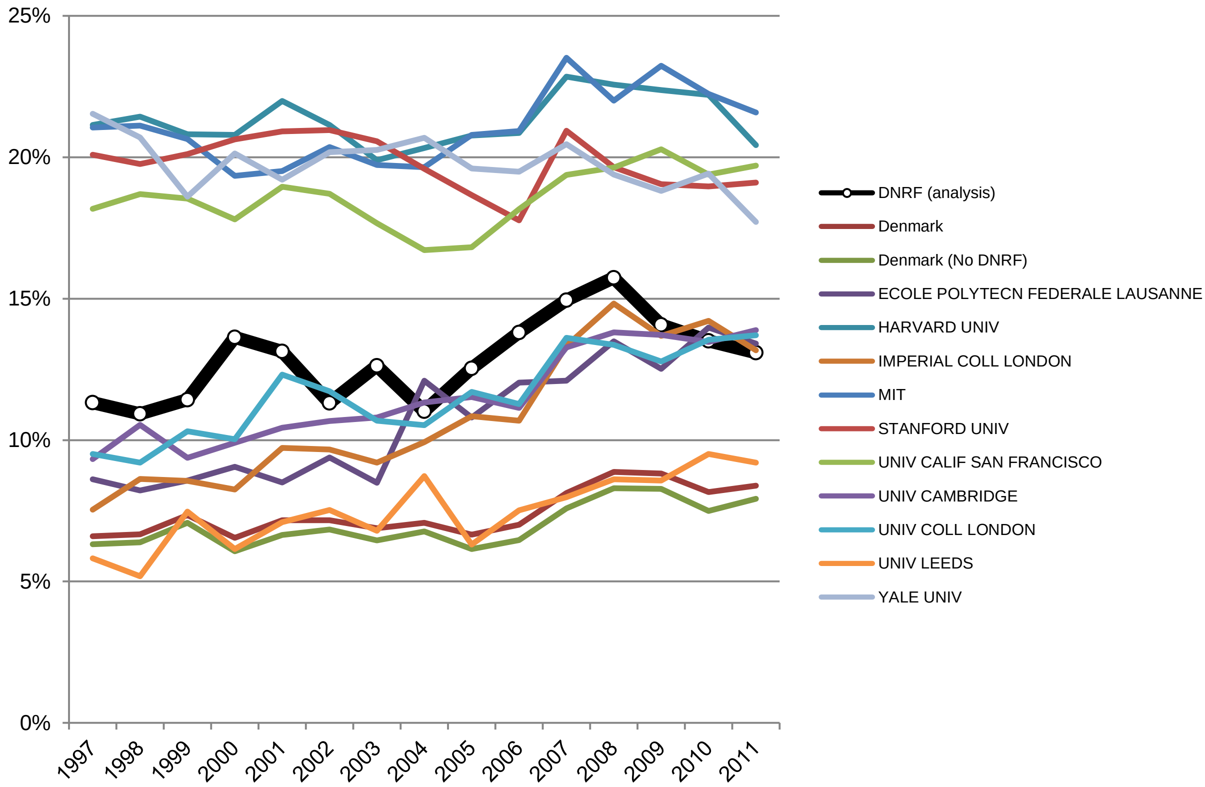



Figure 9.1 shows that the larger units of analysis (e.g. Harvard, Stanford, etc.), have the highest number of publications in high impact journals, while the DNRF (as one of the smallest units) have a much more modest number. For this reason, the analysis in Figure 9.2 below shows the relative shares of publications in 'high prestige' journals. For the whole period combined, the DNRF-publication set has the largest share of publications in 'high prestige' journals compared to the other

European benchmark universities; again American benchmark universities are above DNRF and the European benchmarks. The annual share of DNRF-publications in 'high-prestige' journals varies between 10 to 15% usually somewhere in-between but the pattern is slightly decreasing towards the end of the period. Except for the last three years, the DNRF-publications have the highest annual share among the European benchmark universities. In the last three years the share are almost equal around 13% except for Leeds University at 9%. Also, the annual drop in the share of Danish publications in the 'high prestige' journals varies between .3 and 1.0 percentage points when the DNRF-publications are excluded.

**Figure 9.2: Trend analysis of the relative share of publications in high impact journals over time for all the units.**




## Journal performance according to CWTS's field normalized journal citation indicator: MNJS

In this section we analyse the field normalized impact of the journals in which the different units of analysis has published, denoted by MNJS. If a unit has an MNJS indicator of one, this means that on average the group has published in journals that are cited equally frequently as would be expected based on their field. An MNJS indicator of, for instance, two means that on average a unit has published in journals that are cited twice as frequently as would be expected based on their field citation activity.

Figure 9.3 presents the overall MNJS indicator for the DNRF and Denmark, including and excluding DNRF-publications. Again, a higher publication visibility of the DNRF-publications can be seen. It is also important to mention that the impact of the journals of publication shows an increasing pattern overtime for both the DNRF and the whole country.

**Figure 9.3: MNJS scores for the DNRF-publications and Denmark, including and excluding DNRF-publications.**

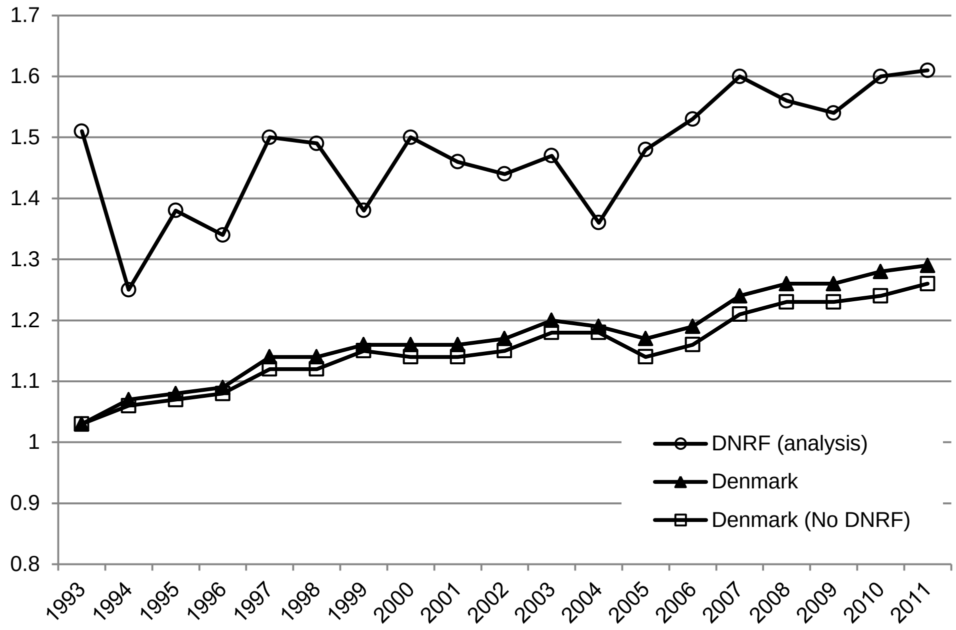
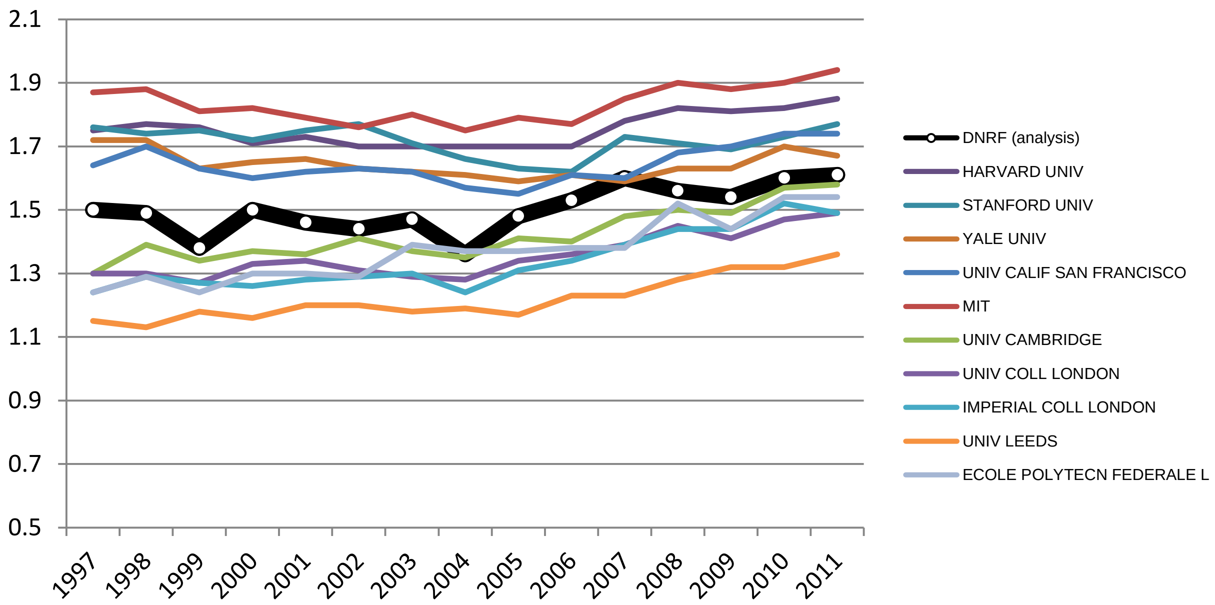



**Figure 9.4: MNJS scores for the DNRF-publications and benchmark universities.**




When comparing the DNRF publication patterns with that of the benchmark units (Figure 9.4), it is again clear how the American benchmark units also have the highest scores in MNJS. It is again interesting to notice that the set of DNRF-publications are published in journals that have a higher field normalized impact as compared to all the other European benchmarks (thus corroborating also the results presented in the previous section). The closest upper benchmark of the DNRF is the

University of California San Francisco, while in the European realm we can mention the University of Cambridge and the Ecole Polytechnique Federale de Lausanne. All in all, the DNRF presents an intermediate position between the American and the European benchmarks, although in the most recent years, the differences between these two main blocks in terms of journal visibility has been reduced.

# 10. SUPPLEMENTARY ANALYSES: ‘Recruitment’ of new top scientists and identification of ‘breakthrough’ research papers

During the work on the main report –presented previously in Chapters 2 to 9 – two extra supplementary analyses were commissioned. These two analyses were carried out subsequent to the main analyses and presented in this chapter. The two analyses concern: 1) DNRF’s ability to ‘recruit’ or attract new scientist which are able to publish highly cited publications within a short time period after their first publication in WoS; and 2) an exploratory bibliometric analysis that seeks to detect ‘breakthrough’ research papers. Notice, the set of DNRF-publications is again the basis for these two analyses, yet the general data set used for the analyses, as well as the benchmark units are different. In both analyses, identification and calculations have been carried out based on the whole WoS database from 1993 to 2009/2011, and in the ‘recruitment’ analysis, benchmark units are countries and not institutions as is the case in the main report. The two analyses are presented below in two sections. Notice, as the development of the two methodologies, as well as the execution of the two analyses have been done subsequent to the writing of the draft report, we present both the methodology and results for the two analyses together.

## The ‘recruitment’ rate of new scientists publishing highly cited publications

In a recent Swedish report (Karlsson & Persson, 2012), the ability of certain countries to recruit new top scientist is estimated by use of a bibliometric approach. Investigating several time periods, the idea in the Swedish report was to estimate a ‘recruitment’ rate for each country, which is supposed to indicate how many new scientists would produce highly cited publications from period to period. The method is not without problems, especially when it comes to identifying researchers. Before 2008, authors and their addresses are in most of the cases not directly linked in the bibliographic data from WoS, which means that name linking and disambiguation becomes difficult and results are therefore to a considerable degree subjected to uncertainty. As the unit of analysis is countries, the approach taken in the Swedish report was basically to severely limit the data set to include only national publications (i.e., publications with authors from the same country), even though this excludes many publications, including potentially highly cited ones (e.g., as mentioned in Chapter 8, it is well-known in bibliometrics that international collaboration brings the highest returns in terms of highly cited publications).

In this analysis, we basically explore the same question as in the Swedish report, namely to estimate the rate at which new top researchers are ‘recruited’. In our case, the unit of analysis is scientists that through their CoEs are funded by DNRF and therefore they function as a proxy for the overall ability of DNRF to ‘recruit’ or produce new top scientist. Compared to the Swedish analysis, we use a substantially different approach centred upon an advanced name disambiguation algorithm developed at CWTS, Leiden University. We compare the ‘recruitment rate’ for DNRF to country benchmarks: Denmark, Finland, Netherlands, Sweden and Switzerland. We use countries as benchmarks in this analysis because country as a data entity in name matching procedures is much more affordable and reliable to handle compared to institutions, such as those used in the other analyses in the report. Notice, due to the link problems between authors and addresses before 2008, numbers presented below should be treated with caution due to the uncertainty of name matching, yet patterns seem robust and informative. We should also emphasize that our methodology is for now explorative.

**Methodology**

In this section we explain the main methodological developments of this analysis, the main definitions used and also the main steps and thresholds considered.

*Basic definitions used in the methodological development*

- Scientists: individual scientist identified through the new author-name disambiguation algorithm at CWTS (Caron & van Eck, 2013).
- Top publication: a publication among the top 10% of the most cited publications in the same field(s). This is calculated following the CWTS methodology of calculating top publications (cf. Waltman & Schreiber, 2012).
- Top or successful scientists: a scientist that has published his or her first top 10% highly cited publication within the 3 years after his or her first publication year.
- First publication year of a scientist: this is the year of the first publication identified in the WoS for an identified scientist.
- Author-affiliation certain linkages: these are linkages between authors and affiliations that are 'certain' or 'true' (i.e. the author can with certainty be linked to a particular affiliation/country in a publication). Thus, certain linkages between authors and countries are the following:
  - First author with the first country in the paper. We can assume that the first author in a paper is in general linked to the first affiliation in a paper (see for example Calero et al, 2006).
  - All authors, where there is only one country represented in the paper (i.e. cases of no international collaboration). Obviously, in a paper where there is only one address or several addresses form the same country we can assume that all the authors of that paper have been affiliated to that country (notice that country of affiliation is determined by institutional addresses and *not* an author's national origin).
  - Reprint author with the reprint country. In WoS the corresponding author of the paper is linked to the corresponding affiliation (and country) (cf. Costas & Iribarren-Maestro, 2007).
  - Direct links of authors and affiliations in WoS extracted from the papers. This information is most reliable from 2008 onwards (cf. Karlsson & Persson, 2012).
  - Most common country (MCC) of a researcher. This is the country among the publications with 'certain linkages' of an author that appears more frequently. In other words, it is the most common country among the countries with which an author has been affiliated. It may happen that sometimes more than one country have the same top frequency for the same author (e.g. an author that has exactly the same number of publications in two countries). In this case, we assign the author to both countries, as in practice it is possible that an author has a double affiliation, or has

been mobile between the two countries, without any of the two countries having a clear majority among his or her papers.

*Main methodological steps*

In this section we describe the main methodological steps developed for the identification of new scientists and their ability to produce highly cited publications over time as a function of their first year of publication. The basic premise is the following: if for every year we count the number of new scientists that started to publish in that particular year and subsequently also those that in the next 3 years (first publication year + 2 more years) published at least one highly cited publication (among the top 10%), we can study the ‘recruitment’ of ‘successful’ new scientists over time within a country and the DNRF-set.

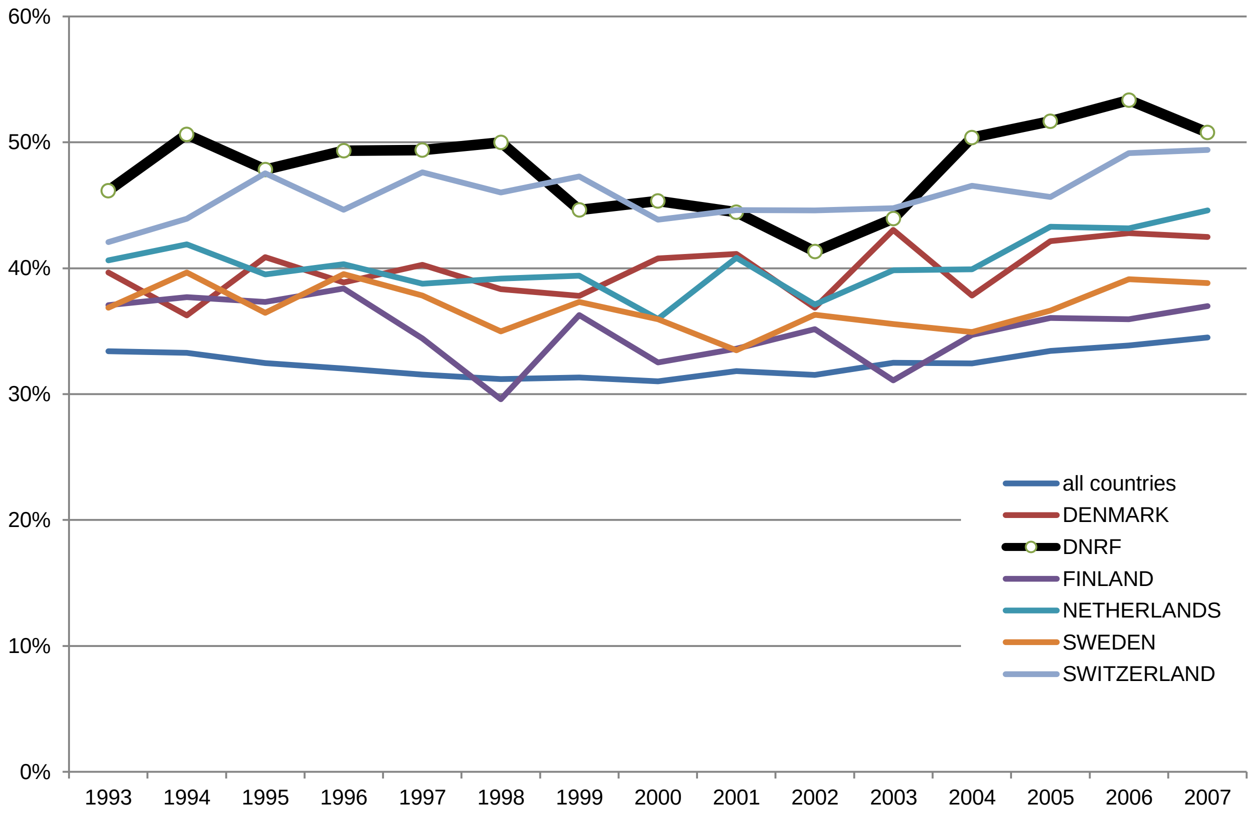
- Step 1. Selection of all publications and authors in WoS during the period 1993-2011. Given the fact that in our main analyses of DNRF we excluded CoEs, from the Humanities, in this step we have considered only publications that are indexed in the Science Citation Index and the Social Sciences Citation Index, thus excluding those publications that belong to the Arts & Humanities Citation Index. This step amounted to more than 90.885,730 author-publication combinations and a total of 22.871,295 unique publications!
- Step 2. Detection of all the scientists active in the previous dataset and their first publication year. This is done based on the recent author-name disambiguation algorithm (Caron & van Eck, 2013) developed at CWTS. We detected a total of 22.377,560 unique scientists. We also identified their first year of publication in WoS. For this we considered the full oeuvre of the identified scientists (i.e. their production in the whole period of our database, ranging from 1980 to 2012), thus it is also possible to detect those scientists that started to publish before the year 1993 and we could exclude them from the analysis and setting homogenous cohorts of scientists. We have also excluded all scientists with less than 5 publications in the whole period, as well as those that started publishing after 2009. A total of 2.128,074 scientists from all over the world meet these criteria and were included for the final analysis.
- Step 3. Calculation of the MCC for all the previously identified scientists. In total 2.072,696 (97%) of all identified scholars got at least one ‘certain’ country or MCC.
- Step 4. Identification of ‘successful’ new scientists as those with at least one highly cited publication within three years of their first publication year in WoS. In total for the whole database, for the period 1993-2009 we have detected a total of 714,152 (34%) of ‘successful’ scientists worldwide.
- Step 5. Identification of DNRF scientists. We designated any scientists with 1) his or her MCC as ‘Denmark’ and 2) more than 10% of his or her publications identified in our previous analyses as belonging to the set of DNRF-publications, as ‘DNRF-scientist’. A total of 13,198 scientists (0.6 % of the total) have ‘Denmark’ as (one of) their MCC and a total of 1,885 (14 % of Danish scholars) can be considered as ‘DNRF-scientists’.
- Step 6. Analysis of the development of new scientists and ‘recruitment’ of new ‘successful’ scientists for Denmark, the DNRF and a selection of benchmark countries (Finland,

Netherlands, Sweden and Switzerland). We also consider all scientists identified worldwide (i.e. the whole database).

**Results**

In Figure 10.1 below we present the 'recruitment' rates for 'successful' new scientists over time for DNRF and the set of benchmark countries; actual numbers for the main results are presented in Appendix V. The numbers in Figure 10.1 should be interpreted such that an annual rate, for example of 50%, means that half (50%) of the new scientists that commenced publishing in that particular year have produced a highly cited publication (top 10%) within three years from this particular starting year. Notice 'producing a highly cited publication' means being an author or co-author of such a publication; the actual contribution to such a publication is not considered. We therefore emphasize that the actual 'recruitment rates' should be interpreted carefully, in as much as they to a large extent reflect collaboration practices. What is more important to notice, are the actual patterns and differences between the units of analysis; here we see some expected trends. First, the 'success rate' for DNRF is the highest overall, closely followed by Switzerland. Yet, in the years from 1998 to 2003 there is a drop for DNRF and here the rate for Switzerland is a slightly higher than DNRF, except for 2000 and 2001. Interestingly, these years corresponds to the years identified as the low points in the impact analyses in the main report (see Chapter 3).

**Figure 10.1: Development in the rate of 'successful' new scientists, where 'successful, means publication of at least one highly cited publication (top 10%) within three years from the scientist's first publication in WoS.**



As the period of the drop is sustained over several years, it cannot be an artefact of annual fluctuations in publication output for DNRF. The drop seems genuine.

In bibliometric evaluations, Switzerland is usually among the top three ranked nations in the world, the others usually being USA, UK and the Netherlands. Denmark is usually on par with the Netherlands, and ranked above the other Nordic countries, such as Sweden and Finland. In that respect, the differences between the countries in this analysis verify the expected pattern. For example, Switzerland is *per se* the best performing country of the five in the analysis and Switzerland is also able to produce or 'recruit' a higher rate of new top scientists compared to the other countries. Hence, it seems that 'recruitment rates' are correlated to impact rank order. Further, if we compare our results to the results in the aforementioned Swedish report (Karlsson & Persson, 2012), which are based on a different approach, the results are actually similar. We see the same rank order between the countries based on the ability to 'recruit' new top scientists. Hence, we therefore consider the results in this analysis to be reliable and also more robust than the Swedish analysis, and we are convinced that the 'success' rate for DNRF with its temporary drop is genuine.

Notice, we have left out 'recruitment' rates from 2008 and 2009 in Figure 10.1 because here we see a sudden rise for all units in rates of 'successful' scientists, but this is an artefact due to a combination of the way we calculate top 10% papers, where tied papers around the $90^{th}$ percentile are weighted (see Chapter 2) and low citation density for the latest years investigated. The latter means less variation and therefore more ties and eventually more units assigned to the top 10% albeit with lower weights. As citation distribution stabilizes over time we see more variation and fewer ties around the $90^{th}$ percentile. Most important, the pattern is basically equal for all units, yet for interpretative purposes we find it most convenient to leave it out as it gives a clearer picture and it does not change the overall conclusions.

Two other things should be kept in mind when interpreting the results. 1) the actual number of scientists for DNRF are considerably lower than the benchmark countries, thus making the rates more vulnerable; and 2) the likelihood for being a co-author on a highly cited publication (i.e., 'successful' scientist) is probably relatively high in the case of DNRF, given the fact that the set of DNRF-publications in general performs at a very high level and the fact that co-authored publications are the norm for all CoEs (see Chapter 6).

With this in mind, we can conclude that the 'success rate' for DNRF is genuine. Overall, DNRF has the highest 'recruitment' rate of the six units analysed. The recruitment rate is approximately 50% for most years analysed; however, there is a minor drop from 1999 to 2003, where the rate drops to approximately 45%. In this short period, the 'recruitment' rate for Switzerland is on par or slightly surpasses the rate of DNRF. Hence, in general one in two scientists affiliated to CoEs funded by DNRF have been associated with at least one highly cited publication within three years of their first identified publication in WoS. Also, for the whole period, 14.3% of the new scientists identified for Denmark are associated with CoEs funded by DNRF and 17.4% of the 'successful' new scientists in Denmark are associated with DNRF.

**Reflections**

As indicated above, the results come out as one would expect for countries and matches those of the Swedish report, even though we apply a very different methodology. Hence, we have no reason to believe that the overall patterns will change if we, for example, changed the thresholds for highly cited papers (for example to top 5%) or the number of highly cited papers needed for a scientist to be considered a ‘top scientists’ (more than one paper in three years). In the Swedish report, the latter threshold is indeed higher, but the publication window is also considerably larger. The results seem robust and scalable. We find that our methodology have some advantages. The analyses are carried on the same conditions for all publications and their authors in the whole database with an advanced author detection algorithm, making it highly robust. Compared to the Swedish report, the approach we take is transparent and simple, with a simple definition of a ‘successful’ scientist and linking of scientists to countries. In fact, 97% of all scientist identified in the analysis (i.e., the whole database) turned out to have an MCC indicator (i.e. most common country). Also, the results are straightforward and replicable also with different thresholds. The methodology also has a number of limitations besides the ones already mentioned above. The data quality is certainly not optimal and the author-identification algorithm is not perfect, although it strives for precision (i.e. that every author is unique although some of his or her publications can be misclassified) and early tests have shown precision and recall values of 95% and 90% respectively.

## Defining and detecting potential 'breakthrough' publications using bibliometric tools: Three explorative approaches

The purpose of this analysis is to explore DNRF's involvement in 'breakthrough research'. Until now, the analyses have focused upon overall impact or impact in major fields and subfields. Using pre-established fields based on journal subject categories is in many ways conflicting with the network of scientific papers, a network that is self-organizing at the paper level and across arbitrary field delineations. Further, knowing that a unit of analysis has high impact in a field does not address what research actually contributes to the high impact. Also, high impact is not necessarily the same as 'breakthrough' research. In this analysis we want to explore the possibilities for detecting 'breakthrough' papers with three slightly different citation analyses.

'Breakthrough' research is obviously a challenging concept to define, operationalize and detect. Numerous approaches can be taken and eventually what can be considered 'breakthrough' research is matter to be decided by peers. Hence, it is foolhardy to believe that a quantitative attempt at detecting 'breakthrough' research with one specific approach can be exhaustive or flawless, this is clearly not case, and we fully acknowledge the limitations of our citation based approach.

Nevertheless, as *one* modest attempt, citation analysis is an interesting approach to explore in this respect. If we assume that in the fields we analyse in this report, research results are mainly reported in international journal articles. If we also assume that within narrower research areas, highly cited publications to a large extent signal impact and use of the content in these papers by the research community, though noise will also be in there, then it would also be reasonable to assume that potential 'breakthrough' research in many instances would be reported in papers that subsequently become highly cited exactly because the research has 'breakthrough' potential. These are the basic assumptions of this analysis and if they are accepted, two major methodological challenges remain: 1) detecting potential 'breakthrough' papers among the set of (extremely) highly cited papers, and 2) establishing an exhaustive network of research areas in which 'breakthrough' papers can be detected. We think the latter is important because analysing potential 'breakthrough' research should commence in the local context of the research area where such knowledge claims are first proposed. We established such a network of research areas by clustering all research papers in the WoS database according to their citation links. Hence we are able to establish clusters of papers at three levels, where at the disaggregate levels papers are clustered because they have similar citation preferences, thus it is assumed that they have common research interests.

We should emphasise that distinguishing between potential 'breakthrough' papers and potential 'breakthrough' research is difficult and also depends on the unit of analysis. A paper can indeed report what eventually turns out to be 'breakthrough' research, but what eventually turns out to be 'breakthrough' research can also be the sum of knowledge claims in a number of papers, where some of them are perhaps not highly cited. In this analysis, we have chosen a simple approach. We assume that the three different citation analyses explored identify potential 'breakthrough' papers given their respective parameters. Subsequently, when we analyse the results, special focus is given to CoEs with several breakthrough papers as we assume that this constitutes a sustained research effort or 'breakthrough' research. Notice, we fully acknowledge that this selection approach only focuses upon a select few research areas disregarding potentially other interesting 'breakthroughs'.

Therefore, this analysis can be seen as substantiating the previous impact analyses, as we further scrutinize the highly cited papers. Notice, while the analysis does not include specific benchmarks, the identification of potential 'breakthrough' papers is done in competition with all the other eligible papers in the WoS-database!

**Defining 'breakthrough' papers**

To precisely define what constitutes a 'breakthrough' paper or 'breakthrough' research is very difficult and eventually turns out to be rather vague. For example, Karlsson & Persson (2012) define breakthrough papers just as those publications in the top 10% of the most cited publications of their fields. If we look for the most common dictionary definition of 'breakthrough' (focusing on those meanings more related with this report) we find the following definitions:

- Oxford dictionary[6]: 'a sudden, dramatic, and important discovery or development'.
- Collins Dictionary[7]: 'a significant development or discovery'.
- Dictionary.com[8]: 'any significant or sudden advance, development, achievement, or increase, as in scientific knowledge or diplomacy, that removes a barrier to progress'.
- The Free dictionary[9]: 'a major achievement or success that permits further progresses.

In general, we can see that there is no straightforward definition of what constitutes 'breakthrough' research and the concepts used to define breakthrough are rather vague and difficult to operationalize. However, based on all the previous definitions we can select a few characteristics that will help us to frame and implement our methodology, these characteristics are the following:

- Potential 'breakthrough' research is reported in journal articles; hence articles are carriers of 'breakthrough' research. The reception of these articles is the determinant of potential 'breakthroughs'. Publications become potential 'breakthrough' papers, whereas a piece of 'breakthrough' research can be reported in one or several papers.
- 'Major achievement or significant and important advance that permits further progress' can be operationalized as publications that are (extremely) highly cited (i.e. they can be considered as 'major achievements' from a bibliometric point of view) and have also had knowledge dissemination and influence within its own field but also in other fields (interdisciplinary spread), thus contributing to further progress within and outside the same field(s).
- 'Sudden or dramatic advance' can be operationalized as publications that are not simple 'followers' of other 'breakthroughs', but they are important publications on their own and therefore a have a distinctive nature compared with previous publications.

[6] http://oxforddictionaries.com
[7] http://www.collinsdictionary.com
[8] http://dictionary.reference.com/
[9] http://www.thefreedictionary.com/breakthrough

Based on the previous concepts we can propose the following definition of 'breakthrough' paper (at least for the purposes of this project): *highly cited paper, with an important spread over its own field(s) and also other fields of science, and it must be a paper that is not a mere follower of other highly cited publication(s) but that it has a genuine relevance on its own*.

In the methodology section we will operationalize this definition. In any case, we acknowledge that this definition of 'breakthroughs' is by no means absolute, other definitions are still possible and valid (e.g. breakthrough papers can also be 'connecting' papers and not necessarily only highly cited ones) and in many cases it would be better to simply talk about remarkable publications. However, in this report, just for language consistency, we will name these remarkable publications as 'breakthroughs'. So, basically our methodology sets out to scrutinize citation patterns, and to distinguish, filter and select between the most highly cited publications in the WoS-database.

**Methodology**

We approach the detection of 'breakthrough' papers from three different perspectives in order to provide different typologies of 'breakthrough' papers. In all three cases we use the new classification of scientific publications developed at CWTS (Waltman & van Eck, 2012). The classification of publications into clusters of highly related publications is the basis for detecting 'breakthrough' papers. This classification system is unique in relation to bibliometric analyses as it is based on direct citation relations between publications and not some arbitrary journal classification like Thompson Reuter's subject categories. The main features of this classification are the following:

- It is a paper-based classification. Publications are classified individually, thus avoiding the limitation of publications that are classified by the topic of the journals they are published in and not by their own content. Besides, the classification is based on articles and reviews only.
- It covers all eligible WoS-publications from 1993-2012, thus fully covering the whole period of this study (1993-2011).
- It is a hierarchical classification with three levels of disaggregation. There are 21 macro-fields that represent main scientific disciplines. These macro-fields contain themselves 784 different meso-fields, and finally we have a micro-classification composed by 21,167 micro-fields. All these levels have been used in our methodology for detecting 'breakthroughs' in one way or another.
- Publications are classified into one single cluster (at any level), thus avoiding the problem of multi-classification of publications and the subsequent multiplicative effects.

It is important to notice that our approach is general and international. We have not been restricted to analysing papers only by one country (e.g. Denmark) or one organization (e.g. the DNRF), although we have studied the presence of these publications for both Denmark and the set of DNRF-publications. Hence, detection of 'breakthrough' papers for DNRF is based upon the whole population of publications covered in WoS during the period 1993-2011, it is competitive and international. Notice, for consistency reasons (i.e. the fact that CoEs from the humanities are excluded from the overall analysis of DNRF) we have focused only on publications from the Science Citation Index and Social Sciences Citation Index (thus excluding the Arts & Humanities Citation Index). Also, for detection of 'breakthrough' papers we exclude reviews as document type and only consider research articles. Based on our previous definition of breakthroughs, 'review' papers can hardly be defended as potential breakthroughs as they mostly condensate and discuss the most recent and important developments and topics in a scientific domain, thus qualifying more as 'followers' than actual breakthroughs under our definition[10].

*CSS method*

All benchmark identification approaches start with a first selection of the most cited publications within each of the 784 meso-fields. However, this selection is not based on the more traditional percentile approach where publications are ranked and top publications are selected from a percentile value within each field as we have done in all other impact analyses in this report (cf. Waltman & Schreiber, 2012). For approach 2 and 3 of this 'breakthrough' analysis, we use the so-called 'Characteristics Scores and Scales' (CSS) method suggested by Schubert et al (1987) and demonstrated in other bibliometric approaches by Ruiz-Castillo (2011). The interesting possibilities of using CSS for research performance analyses are discussed by Glänzel et al (2013). The CSS method focuses on the common characteristics of citation distributions across fields and is based on the principle that citation distributions share some fundamental characteristics and similarities. The CSS method basically consists on the reduction of the original citation distribution to self-adjusting classes by iteratively truncating the distribution to conditional mean values from the low end up to the high end.

In practical terms the method works as follows: taking the distribution of all publications in the WoS classified into the meso-categories, we calculate the mean of the number of citations of the distribution per meso field (m1); then we separate papers above and below the mean, and subsequently for the papers above the mean, we calculate a second mean (m2), again we separate the papers above and below the m2, and to those papers above the m2 citation mean, we calculate a third mean (m3), and finally we again separate these publications above and below m3. As a result we can assign publications to 4 typologies which are described below. In addition to assigning publications to different typologies, as shown by Ruiz-Castillo (2011), this methodology also permits us to characterize the typologies according to the citations they receive. In Appendix VI we present the average values for the four typologies, as well as the proportion of citations they receive. As can be seen, there is a remarkable regularity across fields of science and across our meso-fields. This is very useful for our purposes in this study as it allows us to somehow to characterize the 'success' and 'dissemination' of the impact of these publications. The four typologies can be characterized as follows:

[10] However, they have been included in our CSS methodology, thus making the top classes of this methodology more exigent from a citation reception point of view.

- Type 1. *Lowly cited publications*: those that have an impact below the average of the entire field (m1). They are the vast majority of publications in every field representing around 74% of all the publications and accounting for less than 22% of all the citations.
- Type 2. *Moderately cited publications*: those that have an impact above the average of the entire field (m1) but below the second mean (m2). These publications represent less than 20% of all the publications in their fields and receive 32% of all the citations in the field.
- Type 3. *Highly cited publications*: these are publications that have an impact higher than m2 but below m3. They are around 5% of all the publications within each meso-field and receive more than 21% of the citations in their respective fields.
- Type 4. *Outstanding publications*. These publications represent barely 2% of all the publications in every meso-field, but they alone receive around 25% of all the citations in their meso-fields. In other words, these are the 2% most cited publications of every field and one in four citations given in their meso-fields goes to them.

Based on this methodology, 16.250,505 publications covered by WoS from 1993-2012 and that has a meso-field in the CWTS classification[11] have been classified (in this case both articles and reviews). A total of 314,944 (1.9%) publications belong to type 4 (i.e. outstanding publications), of which 263,148 are of the document type 'article' (1.7% of all articles in the period). These 263,148 publications have been considered as 'potential breakthroughs'.

*Filtering 'followers'*

As discussed previously in our definition of 'breakthrough' papers, being highly cited is in itself not sufficient to be considered a breakthrough, because publications should not be *a mere follower of other highly cited publication(s)* – it must have *a genuine relevance on its own*! In this sense, it is not at all uncommon that highly cited publications are so just because they have followed the steps of a previous breakthrough (or breakthroughs) and somehow they profit from the 'spell' (i.e. innovativeness, novelty, relevance, etc.) of those previous publications. To operationalize this filtering of the 'followers' we have performed the following steps:

- Identification of all pairs of potential breakthrough papers. Basically, we have identified potential breakthroughs citing another potential breakthrough(s). If we find such linkages, we label the citing breakthrough as B2 and the cited breakthrough as B1. Thus B2 papers are *potential* 'followers'.
- We then analyse the papers that cite B2 and check if they also cite B1, if so, we count these papers as double citers of B1 and B2.
- Finally, for B2 publications, we count how many of their citing papers that either simultaneously cites both B1 and B2 or only B2. Subsequently we enforce a threshold to designate 'followers'. Thus, for every potential 'follower' (i.e. B2 papers) *not* to be finally designated as a 'follower', the paper must have 70% or more of its citations alone (i.e. not being co-cited with B1 in more than 30% of its citations). The main idea behind this

[11] A total of 15,498,978 publications are classified as document type 'article' in the Web of Science in the same period.

threshold is that a breakthrough should not benefit too much from the spell of a previous breakthrough, as it should have a genuine impact on its own.

We have applied this filter to the 263,148 outstanding articles previously detected and 179,347 passed the filter of the followers being considered as potential breakthrough candidates (they are used later on in Approach 2). The 179,347 breakthrough papers correspond roughly to 1% of the publications in the period 1993 to 2011. We now describe the three different perspectives or approaches we have applied to detect breakthrough papers – including DNRF-breakthrough papers.

*Approach 1: 'Breakthrough' papers based on the micro-classification*

The first approach is very simple but also extremely exclusive. It is based on the idea that the most cited paper of every micro-field can most likely be considered a 'breakthrough' paper because it has the highest impact in it its micro-domain. This is a very restrictive definition of a 'breakthrough' paper, because only *one* (or occasionally several) papers pass this filter. In fact, only 21,670 publications pass this filter as breakthroughs. For the time being, we ignore other potential dimensions of breakthrough such as knowledge diffusion and use this simple restrictive approach as a first general benchmark on what we can expect from the two other approaches.

*Approach 2: 'Breakthrough' papers detected through the 'Characteristics Scores and Scales' (CSS) method and filtering of 'followers'*

This approach is based on the 179,349 publications that have passed the filter of the 'followers'. We have considered all these publications as breakthroughs and have analysed the presence of DNRF-publications in this set.

*Approach 3: 'Breakthrough' papers detected through the 'Characteristics Scores and Scales' (CSS) method, filtering of 'followers' and selecting those that have an impact in other different macro fields (above the average of all the breakthrough candidates)*

Based on the 179,349 publications considered for Approach 2, we have included a new filter for determining breakthroughs. In this case, we introduce a multidisciplinary dimension in the delineation of breakthroughs assuming that substantial breakthroughs also have impact beyond their own micro- and meso-domains (i.e. they have an important outreach or knowledge diffusion to other major fields that in return cites these 'foreign' papers). To operationalize this idea, we have taken a relatively simple approach composed by the following steps:

- Taking all the citers of the 179,349 publications previously filtered, we counted the number of different macro-categories (i.e. a total of 21) from which they have received at least one citation.
- We then calculate the average of different external macro-fields where the breakthroughs of every meso-category have had some impact.
- Thus, based on the previous values, we consider a 'breakthrough' those publications within the same meso-category with an impact outside their own macro-field higher than the average of all the potential breakthroughs in the same meso-category.

Therefore, a 'breakthrough' according to this third approach is potential breakthrough papers that have an impact in more macro-categories than an average potential breakthrough within the same meso-category. A final set of 59,617 articles can be considered as 'breakthroughs' according to this approach, approximately 0.38% of all the articles in the time period analysed.

**Comparison**

In order to develop our methodology and to have some sort of yardstick to compare the results of the three bibliometric approaches with, we were provided with a sample of eight CoEs, considered to have produced 'breakthrough' research. Each CoE made a small description of their scientific contribution and indicated a number of papers they considered important. While it is certainly valuable to have such a list, we cannot rule out that the selection of CoEs and/or the designation of important papers are influenced by bibliometric indicators (e.g., the three highest performing CoEs in the DNRF-publication set is included), and if so, there is a selection bias. Further, among the CoEs chosen, some indicated important papers that were published outside the period analysed here, one CoE indicated important papers that were not included in the DNRF-publication set because the publication year in question was missing in the original publication list; and finally one of the eight sampled CoE is still active and for this analysis has only been active in 4-5 years. The latter fact significantly lowers the probability of detecting 'breakthrough' papers. Given the present analysis' focus upon detecting potential 'breakthrough' papers among the set of extremely highly cited publications, we should expect that papers from some of the highest performing CoEs will turn up, whereas the challenge will be further down the impact strata.

**Results**

In this section we present the main results of the three different approaches. As stated above, we mainly focus upon overall findings for DNRF, as well as CoEs with several potential 'breakthrough' papers. Notice, CoEs are kept anonymous. Based on the delineations of breakthroughs previously specified in approaches 1, 2 and 3, we study the presence of these types of publications within the general articles produced by the DNRF. It is important to mention here that the DNRF has produced a total 10,803 articles in the period of analysis. We present the identified 'breakthrough' papers in three science maps. It is the same base map of the 784 meso-fields that is used in the three analyses, where the micro-fields are positioned in relation to each other according to citation links. In these maps, we indicate potential 'breakthrough' papers from DNRF as numbers placed on top of the micro-field (i.e. numbers are the actual number of DNRF 'breakthrough' papers in the micro-field).

*Results of approach 1: 'Breakthrough' papers based on the micro-classification*

Approach 1 is the most simple but also the most restrictive of the three approaches. A 'breakthrough' paper is here defined as the paper with highest the impact in the micro-field. Such a paper is considered to have tremendous importance or constitution for the research area. In all, 21, 167 micro-fields have been established and 21,670 publications pass the filer as 'breakthroughs' (i.e. the number is higher than the number of micro-fields because there are ties among the highest cited papers in particular fields). Of these 21,670 potential 'breakthrough' papers, 0.15% or 32 papers come from the set of DNRF-publications. Notice, the set of DNRF-publications constitutes only 0.07% of all publications analysed, thus there is an overrepresentation of potential 'breakthrough' papers from the set of DNRF-publications. Also, the 32 potential breakthrough' papers constitute 0.3% of the papers in the DNRF-publication set.

**Figure 10.2: Approach 1 - distribution of 'breakthrough' papers among CoEs.**

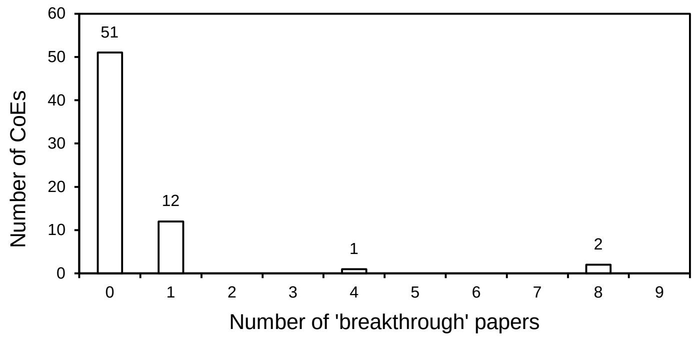



Figure 10.2 shows the distribution of 'breakthrough' papers among the 66 CoEs. The distribution shows the characteristic skewness found in scientometrics: three CoEs accounts for 20 of the potential 'breakthrough' papers. Whereas 12 CoEs have one 'breakthrough' paper and 51 CoEs have no 'breakthrough' papers according to this definition.

**Figure 10.3: Approach 1 – map of 784 meso-fields where potential DNRF-'breakthrough' papers are indicated with numbers on top of fields.**

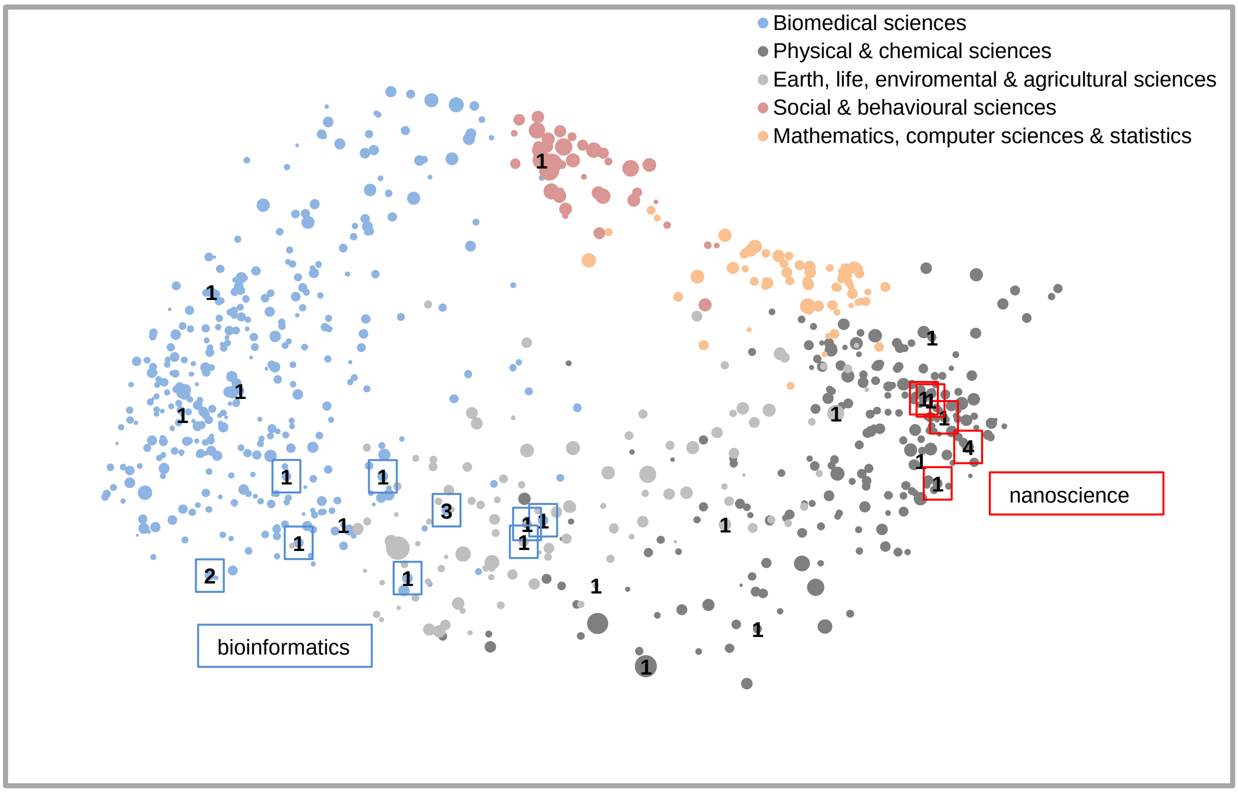


Figure 10.3 above shows the science map where the potential ‘breakthrough’ papers are marked with a number on top of the meso-fields (circles). Notice, the ‘breakthroughs’ are detected in the micro-fields and a meso-field is constituted by several micro-fields. We have identified the major research fields with colours and the size of circles indicates the relative size of the meso-fields.

As stated above, three CoEs account for 20 of the potential ‘breakthrough’ papers. As indicated on the map, 12 of them concern ‘breakthrough’ research in bioinformatics and eight of the papers what today would call nanoscience.

*Results of approach 2: ‘Breakthrough’ papers detected through the ‘Characteristics Scores and Scales’ (CSS) method and filtering of ‘followers’*

Approach 2 is based on the CSS method where so-called ‘followers’ are removed by filtering. Even though, the approach only considers the 2% most cited papers, it is the least restrictive approach of the three we explore in this analysis. In total, 179,349 publications from 1993-2011 have been detected as potential ‘breakthroughs’, of these 241 come from the set of DNRF-publications, this corresponds to 0.13% of all potential ‘breakthrough’ papers defined by approach 2 (and 2.2% of the papers in the set of DNRF-publications).

**Figure 10.4: Approach 2 - distribution of ‘breakthrough’ papers among CoEs.**

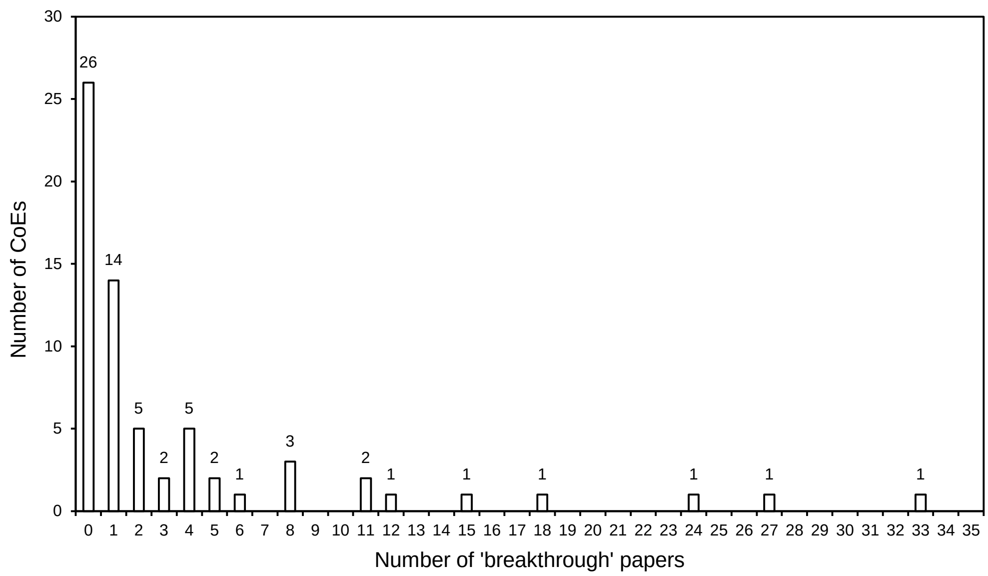
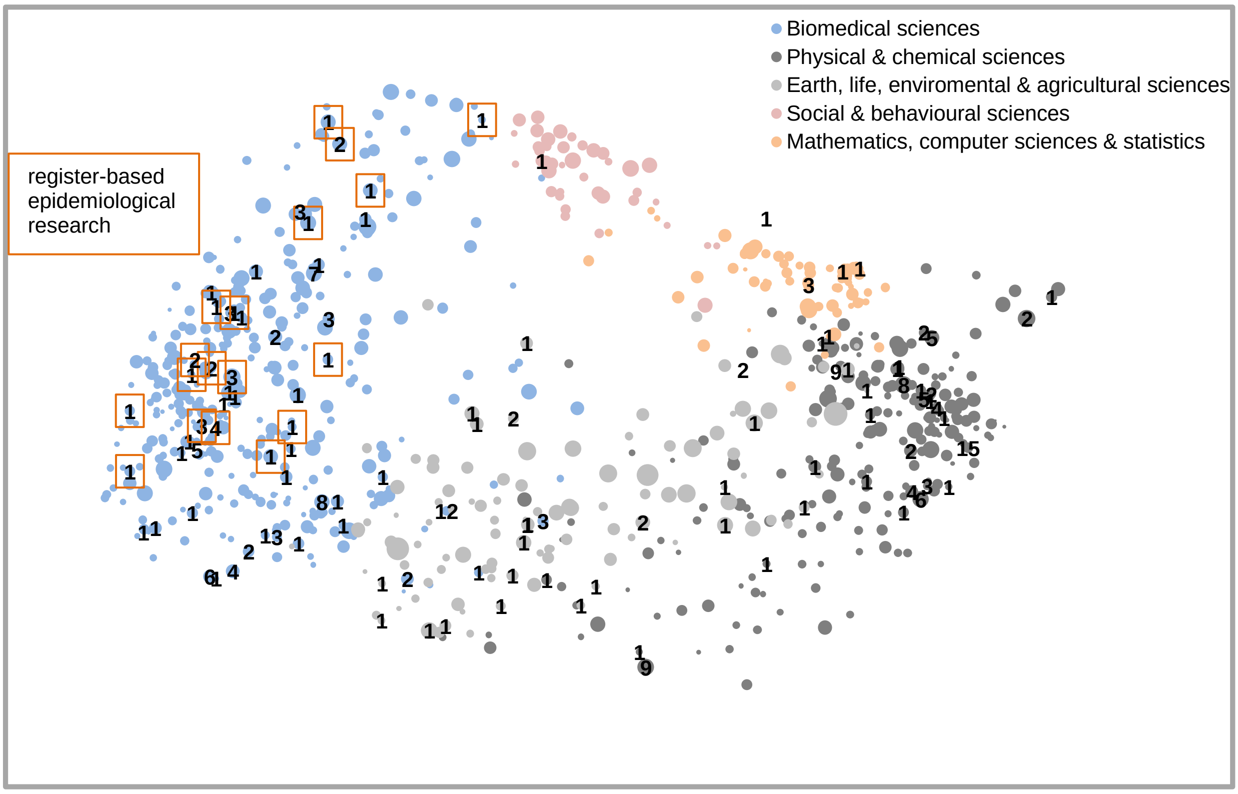



Figure 10.4 shows the distribution of ‘breakthrough’ papers among the 66 CoEs. Again the distribution is skewed, for slightly more than one third of the CoEs (26), no potential ‘breakthrough’ papers are detected. Nevertheless, as expected many CoEs (40) now have at least one potential ‘breakthrough’ paper according to this approach. It is also expected that the three CoEs that together accounted for 62% of the potential ‘breakthrough’ papers in approach 1 also will be highly visible with this approach, given its lesser restrictions. Together they account for 31% of the potential ‘breakthrough’ papers in this approach. However, there are new candidates emerging

from this approach. The second highest number of potential ‘breakthrough’ publications, 27, is from a CoE that had one potential ‘breakthrough’ publication in approach 1.

Figure 10.5 below shows the distribution of potential ‘breakthrough’ papers identified by approach 2 over the 784 meso-fields. In this map we have marked the ‘newcomer’ which concerns register-based epidemiological research. Bioinformatics and nanoscience papers can be located in the same areas or close by as in the previous map in Figure 10.3. What is not indicated are potential ‘breakthrough’ papers in areas such as catalysis, metal structures, as well as muscle and sensory-motor research.

**Figure 10.5: Approach 2 – map of 784 meso-fields where potential DNRF-‘breakthrough’ papers are indicated with numbers on top of fields.**




*Results of approach 3: ‘Breakthrough’ papers detected through the ‘Characteristics Scores and Scales’ (CSS) method, filtering of ‘followers’ and selecting those that have an impact in other different macro fields*

The third and final approach is comparable to approach 2 since it is also based on the CSS method where ‘followers’ are removed. But the approach is much more restrictive because it further delimits the set of potential ‘breakthroughs’ into only those publications that have an impact in more macro-categories than an average potential ‘breakthrough’ paper within the same meso-category. Hence, we try to reflect the potential knowledge diffusion of the potential ‘breakthroughs’. A total of 59,617 articles are considered potential ‘breakthroughs’ according to this approach. Of these, 0.16% or 97 papers come from the set of DNRF-publications and this corresponds to approximately 1% of the DNRF-publications.

Figure 10.6 below shows the distribution of ‘breakthrough’ papers among the 66 CoEs and Figure 10.7 below shows the distribution of potential ‘breakthrough’ papers identified by approach 3 over the 784 meso-fields.

**Figure 10.6: Approach 2 - distribution of ‘breakthrough’ papers among CoEs.**

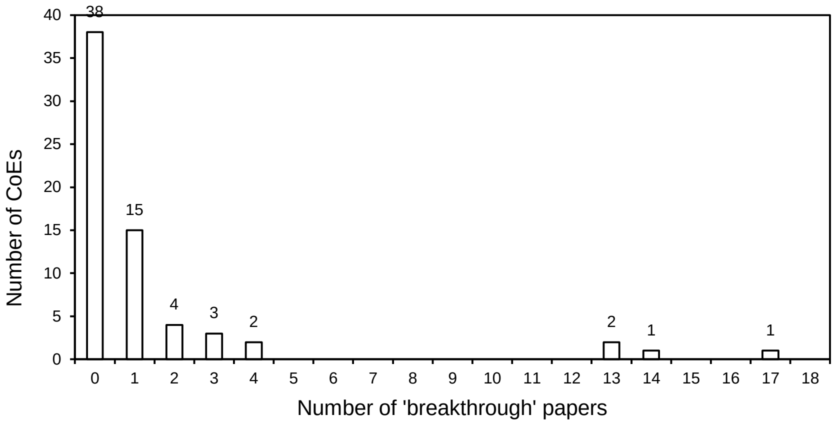
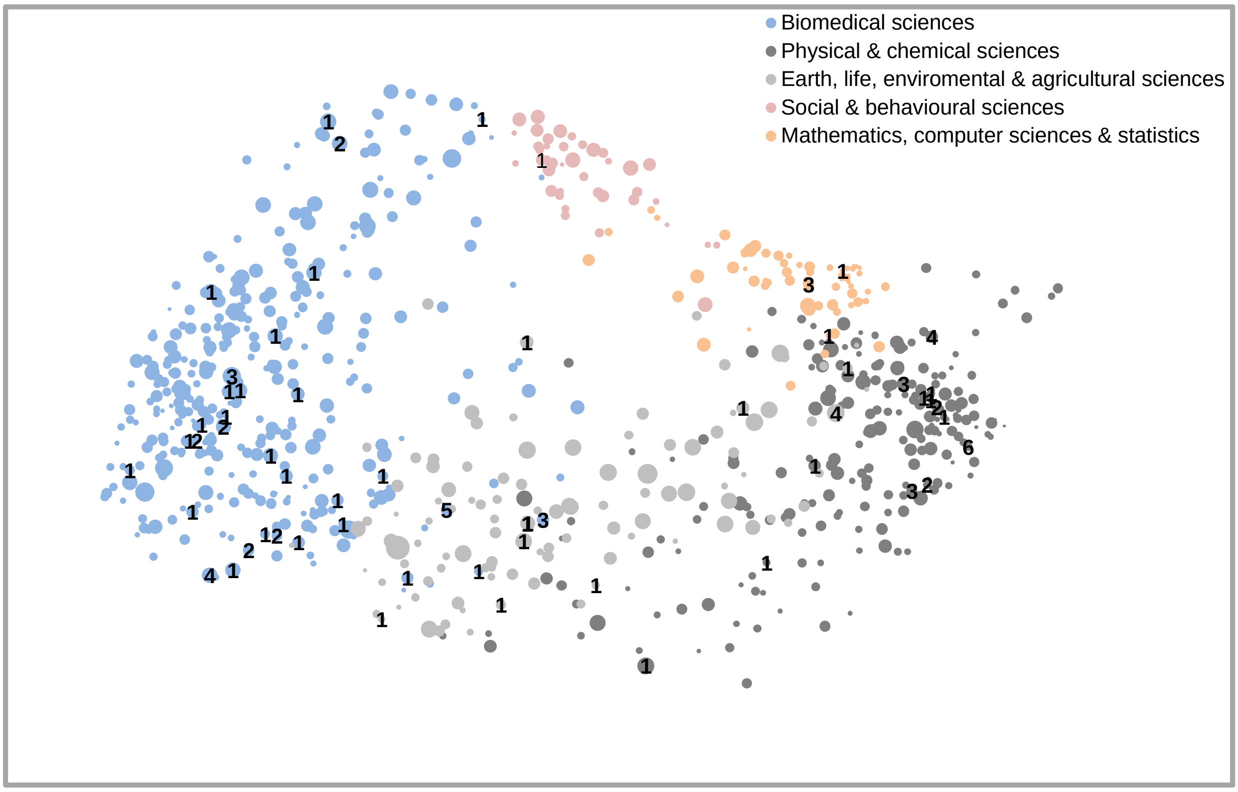



**Figure 10.7: Approach 3 – map of 784 meso-fields where potential DNRF-‘breakthrough’ papers are indicated with numbers on top of fields.**



Again in Figure 10.6 we see the familiar skewed distribution. Not surprisingly given the requirements of interdisciplinary citations from other macro-categories, the number of potential ‘breakthrough’ papers is markedly reduced compared to approach 2 and the number of CoEs where no potential ‘breakthrough’ papers could be identified has increased considerably to 57% of all CoEs. It is also noteworthy that the number of potential ‘breakthrough’ papers is significantly higher for four CoEs compared to the 62 other CoEs. CoEs ranked 5-6 have four potential ‘breakthrough’ papers, whereas the CoEs ranked 3-4 have 13 ‘breakthrough’ papers each!

In Figure 10.7, the four CoEs together account for 59% of the potential ‘breakthrough’ papers identified with approach 3. Not surprisingly it is the same CoEs and research areas as depicted in the other two maps: bioinformatics, nanoscience and register-based epidemiological research.

**Summary, comparison and reflections**

It is very important to emphasise that we have only focused on the most significant signals in the DNRF-publication set. More detailed analyses will indeed show potential ‘breakthrough’ papers for a considerable number of the CoEs. While it is not entirely surprising that the major findings, i.e. the same three to four research areas and CoEs, recur in the results of the three approaches, after all the approaches are variations over the same idea, it is noticeable that these areas come out so strong. Three CoEs have several highest cited publications across the micro-fields. They all have numerous potential ‘breakthroughs’ among the highest cited publications and a considerable number of their potential ‘breakthrough’ papers receive, more than average, citations from other macro-categories. The latter indicating the widespread diffusion and use of this particular research.

If we compare the results to the eight CoE examples, then we can conclude that potential ‘breakthrough’ papers are identified for four CoEs in all three approaches; potential ‘breakthrough’ papers are identified for five CoEs in two or more of the approaches, and potential ‘breakthrough’ papers are identified for six CoEs in one or more of the approaches. Consequently, for two examples we did not identify any potential ‘breakthrough’ papers. Of the two examples, one is still active and its publication activity for the period under investigation is limited, hence not identifying potential ‘breakthrough’ papers in this case may be an effect of time. This explanation does not hold for the other case. Here the funding of CoE expired within the period under investigation. Also, potential ‘breakthrough’ papers are identified for several other CoEs funded at the same time. Hence, we must conclude that given the three approaches we explore in this analysis, no papers from this particular CoE qualify as a potential ‘breakthrough’ papers.

Finally, the three clearly highest ranked CoEs in the PPtop10% impact analysis in the main report are also the three most prominent in this ‘breakthrough’ analysis: bioinformatics and nanoscience.

To conclude, we have tried empirically to detect potential ‘breakthrough’ papers and thus ‘breakthrough’ research. As we have argued the approach has numerous limitations. Particularly we like to emphasise that we are only able to detect strong signals through citation analysis. As this signal becomes weaker we are not able to detect it and given our definition and operationalization not able to identify potential ‘breakthrough’ papers. On the other hand, if the signal turns out to be strong, we will - other things equal - detect it and this we have done in this analysis. Hence, all papers have been treated equally. Yet something will obviously go undetected. Other thresholds and definitions could have been applied and so forth. However, within its limits, we have delineated our approach with reasoned arguments. The methodology is simple and replicable. There is consistency among the approaches and the results, especially among the strongest signals, and the results are in line with the eight examples given to us for validation. Indeed, one could argue that the approaches are too restrictive, but identifying 'important' papers which are located lower down in the citation distributions is very difficult as signals get nosier. Finally, we should state that there may be a time-lag effect in our approach. Citation analyses are per definition retrospective. Of the 26 CoEs where we did not detect any potential 'breakthrough' papers, 19 were still active during the period under investigation. Hence, it is reasonable to suggest that some perhaps most of them will eventually produce potential 'breakthroughs', it just takes some time for them to be visible.

# 11. SOME LIMITATIONS AND CAVEATS

In this brief chapter we summarize in bullets some of the caveats and limitations of the present analysis which should be considered in relation to the interpretation and extrapolation of the results.

- First, there are the 'usual suspects' in bibliometric analyses. The set of publications are limited to the universe of WoS. The data set is therefore limited to mainly English language journals articles, or rather those indexed by WoS. Therefore the data set is also limited to the fields where the main publication activity is in (international) journals, thus due to low coverage, some fields are excluded from the analysis. Obviously, this means that only a subset, albeit a large one, of DNRF-funded CoEs have been included in the analysis and of those included; only the international journal publications indexed in WoS have been the basis for the analysis. With this limitation in mind, it is reasonable to claim that for most, if not all, of the CoEs included, international journals are the main publication outlet and thus a good indicator for scientific performance. Another 'usual suspect' is the validity of the indicators used. No doubt, we have applied the best indicators available, but in order to interpret them in a meaningful way, one needs to accept the assumption that they measure impact and that impact somehow reflects the importance of literature use in the scientific community, and that this eventually indicates something about the usefulness or importance of knowledge claims purported in the publications.

- Publication lists from the individual CoEs originally used for reporting of activities to the DNRF are the basis for extraction of potentially eligible journal publications. The publication lists are not complete and in some instances deficiencies have been detected. In 4 publication lists some publications have been marked - for various reasons - as a result remote from the DNRF funding. In at least 10 publication lists there are gaps between publication years, usually one or two years of publications may be missing in the lists. Empirical analyses of publications from 2008 onwards indicate that some publications credited to the CoEs and thus implicitly to DNRF, acknowledge other Danish funding agencies than DNRF in the funding acknowledgments. Knowing that some publications may be false positives, we decided to treat all potentially eligible journal publications on the lists in the same way by including them all in the analysis. We have done this for several reasons. First of all they appear in the list and thus are used to report activity back to the DNRF. Second, in the 4 cases where some publications are actually marked, they are so from widely different reasons. Third, we do now to what extent such 'periphery' publications are present in the other publication lists. Fourth, clearly some publications are missing from the publication list; some due to the gaps mentioned above and others due to errors in reporting or errors by us in detecting such publications. Consequently, it not realistic in the present analysis to treat each publication list individually in manner so that only publications that are a direct result of DNRF-funding are selected and also ensures that missing publications are retrieved. It is also not necessary as the analyses are on the aggregate DNRF-level, where the individual CoEs' publications are treated as one set. This publication set (some 11600) is sufficiently large to ensure robust indicators (especially PPtop10%), hence false positives or missing publications will not influence the overall conclusions at the DNRF-level in any significant way. For example, if we remove all the publications marked as somehow ‘periphery’ in the 4 publication lists, even though it is questionable whether all four indications should lead to exclusion, the overall PPtop10% for the whole period 1993-2011 goes from 20.4% to 20.9%. The change in the example is minimal. But more important, it is unclear what should be excluded and obviously what should be included as something is also missing. Therefore we consider the approach taken robust on the aggregate level of the DNRF-publication set. Notice, however, that at the disaggregate level of CoE, the indicators become more vulnerable due to missing publications or false positives, especially where publication numbers are already low. We have therefore also stressed that publication output reported for individual CoE should be treated with caution and not be interpreted as actual publication activity.

- The benchmark units used in the present analysis are not ideal. The units of analyses are different. The DNRF-set of publications are defined according to funding, whereas the two benchmark units are geographically and organizationally defined. Ideally, the set of DNRF-publications should have been compared to sets of publications coming from similar excellence funding initiatives. Also, measuring the main effect of being funded by DNRF, requires that publications from rejected applicants are investigated. In this respect, the different sizes of the units of analyses should also be mentioned (i.e., size is defined according to publication output). If possible, units of correspondingly similar size should be compared in bibliometric studies. It is generally so that with larger units, be it institutions or fields, indicator values will tend to move closer towards the reference value, a sort of ‘regression towards the mean’. However, some of the highest performing units, for example Harvard University, manage to both have exceptionally high impact and a very large publication output. We simply cannot know whether the DNRF-set of publications would perform at the same level if the set was considerably larger. In effect, one has to consider that for example Harvard University produces some 14 times more publications than the DNRF-set and still manage to outperform the DNRF-set in most fields. Nevertheless, using these benchmarks, still gives an impression of the overall performance of the set of DNRF-publications.

- Treating the set of DNRF-publications as a unit of analysis is unusual and not without problems. Usually publications are assigned to researchers, research groups, departments, institutions or countries, well-defined discrete units. Assigning publications to a funding institution is somewhat fuzzier as publications are a result of many influences and often several funding organizations, and most importantly, determining to what extent a publication is a result of a certain funding can be elusive.

- For the time being, the question whether the effect sizes of the annual drops in the Danish PPtop10% indicator when removing the DNRF-publications can be considered an important or large drop remains unanswered; we simply do not have anything to compare the results with. Statistical significance tests are not an appropriate solution. Our conclusion that the drops are notable is based on our impression and experience with similar data analyses.

- While we generally think the ‘recruitment’ analysis is sound and the results robust, it is important to emphasize that the automatic detection of scientists are subjected to some uncertainty and that the numbers produced should therefore be treated with some caution.

- It is very important to emphasize that the ‘breakthrough’ analysis is exploratory and given its assumptions, chosen definitions and operationalization has several limitations and is certainly not flawless. Refined citation analyses are only able to detect the strongest signals. As signals become weaker we are not able to detect them and given our definition and operationalization therefore not able to identify potential ‘breakthrough’ papers. However, within its limits, we argue that the approach is interesting; we have delineated our approach with reasoned arguments, it is simple and replicable. If signals turn out to be strong our approach will detect them and this we have done in this analysis. Hence, all papers in the WoS databse have been treated equally. Yet something will obviously go undetected. Other thresholds and definitions could have been applied and so forth.

# 12. CONCLUSIONS

In this final chapter we briefly summaries the main conclusions that can be derived from the study, for more details see the summary at the beginning of the report.

The main conclusion of this report is that the DNRF-publications overall perform at a very high-level comparable to the absolute highest-performing universities in Europe and often slightly better. The DNRF-publications contributes notably to the overall Danish impact given its relative size. However, there are annual fluctuations and marked variations in performance between fields and subfields and between individual CoEs. Especially, noteworthy is the performance in the category 'multidisciplinary journals', here the DNRF-publications performs at the same level as the highest ranking universities in the world.

The following bullets highlight more specific conclusions and recommendations.

- DNRF-supported publications constitute around 7% of all Danish WoS-journal publications in the period 1993-2011. These publications gather around 10% of all Danish citations (9% of all normalized citations).
- Excluding DNRF-supported publications from the set of Danish publications generally results in a decrease of the PPtop10% and MNCS indicators. For the annual PPtop10% indicator, the decrease is on average .5 percentage points, however, in later periods this number has been rising.
- In most impact analyses, the set of DNRF-supported publications is ranked in-between the American and European benchmark universities. In general American benchmark universities have a higher performance than the set of DNRF-publications, albeit the latter have generally a higher performance than the European benchmark universities. Notice, there are variations on the field and subfield levels.
- In general, the set of DNRF-supported publications are especially comparable to the Ecole Polytechnique Federale de Lausanne (the highest ranked university in Europe according to Leiden Ranking 2013/2014), both in relation to size and impact.
- The DNRF produces slightly more than 20% of top 10% highly cited publications, showing an increasing pattern over time. The same pattern can also be observed for the field normalized impact (MNCS) of the publications. In this sense, the set of DNRF- publications have an impact well above the international level (>1.5 in MNCS).
- The set of DNRF-supported publications constitutes around the 11% of all highly cited Danish publications (top 10% publications).
- In all fields and subfields the set of DNRF-supported publications always perform above the international level and for all fields, except the 'social and behavioural sciences', well above the national Danish level.
- It is remarkable that DNRF-supported publications perform at the same level as Stanford University and MIT when it comes to the 'multidisciplinary journals' category (e.g. journals such as Nature, Science or PNAS).

- The PPtop10% indicator in the ‘natural sciences’ ranks the DNRF-set among the American universities just above the University of California, San Francisco and the gap to the European universities is noticeable.
- Compared to the benchmark universities in the ‘medical and life sciences’, the DNRF-set ranks the lowest except for Leeds University according to the PPtop10% indicator and one rank higher (slightly above Imperial College London) according to the MNCS indicator. ‘Medical and Life Sciences’ is the field that contains most DNRF-publications when the period under investigation is combined.
- The general publication output has risen over the years in all subfields. The largest subfields are ‘physics and materials science’, ‘life sciences and biology’, ‘chemistry’, ‘clinical medicine’ and ‘biomedicine’.
- The contribution to the Danish PPtop10% indicator at the level of subfields shows more variation over the period. With varying degrees of contribution to the national indicator in different periods, subfields such ‘astronomy and astrophysics’, ‘physics and materials science’, ‘chemistry’ and ‘geosciences’ show sustained considerable contributions. In later periods, ‘life science and biology’ does the same.
- Considerable variation in publication output and impact was found among the 66 individual CoEs analyzed. Some CoEs perform extremely well, but some 12-13% of the CoEs performs at or slightly below the international level. Grouping CoEs according to their granting year reveals that the median level of performance is higher for CoEs funded in the later periods 2005 and 2007.
- In terms of collaboration, around 58% of the DNRF-publications come from international collaboration. Noticeable, DNRF-publications with no collaboration has a remarkably high impact, comparable to the highest ranked American benchmark universities.
- In a recent subsample it was found that in 47% of the DNRF-publications, other Danish funding agencies were also acknowledged.
- Regarding publication journals, it is notable that more than 10% of the publications supported by the DNRF are published in high prestige journals, showing an increasing pattern over time. DNRF publications are published in relatively more prestigious journals compared to European benchmark universities.
- In general, DNRF has the highest ‘recruitment’ rate of ‘successful’ new scientists of the six units analysed. The ‘recruitment’ rate is approximately 50% for most years analysed; however, there is a minor drop from 1999 to 2003, where the rate drops to approximately 45%. Hence, approximately one in two scientists affiliated to CoEs funded by DNRF has been associated with at least one highly cited publication within three years of their first identified publication in WoS. Also, for the whole period, 14.3% of the new scientists identified for Denmark are associated with CoEs funded by DNRF and 17.4% of the ‘successful’ new scientists in Denmark are associated with DNRF.
- In terms of detecting ‘breakthrough’ papers, given the size of the DNRF-supported publication set, more potential ‘breakthrough’ papers than expected are detected in the DNRF-publication set for all three detection-approaches. The distribution of potential

‘breakthrough’ papers is heavily skewed over CoEs and this is also the case for all three detection-approaches. Three research topics are clearly the most visible among the detected ‘breakthrough’ papers: bioinformatics, nanoscience and register-based epidemiological research. Several other research topics are represented with potential ‘breakthrough’ papers, though not nearly as visible as the three aforementioned topics. Also, the three highest ranked CoEs in the PPtop10% impact analysis in the main report are also the three most prominent in this ‘breakthrough’ analysis (i.e., bioinformatics and nanoscience).

- As discussed in the methodology, one of the main difficulties in this study was the data collection, which has been based on an exhaustive combination of manual and automatic procedures. This makes the analysis a quite unique study of a funding research organization. A potential recommendation for the DNRF is to develop a well-structured database containing all the information on their funded research and supported publications. This would ensure efficient analytical possibilities in the future and a higher capability in monitoring its own supported production and impact.

# REFERENCES


Calero, C.; Buter, R.; Cabello Valdés, C. & Noyons, E. (2006). How to identify research groups using publication analysis: an example in the field of nanotechnology. *Scientometrics*, 66(2): 365-376.

Caron, E.C.M. & van Eck, N.J. (2013). Large scale author name disambiguation using rule-based scoring and clustering. *Forthcoming*.

Costas, R. & Iribarren-Maestro, I. (2007). Variations in content and format of ISI databases in their different versions: the case of the Science Citation Index in CD-ROM and theWEb of Science. *Scientometrics*, 72(2): 167-183.

Costas, R., van Leeuwen, T.N. (2012). Approaching the “Reward Triangle”: general analysis of the presence of Funding Acknowledgments and “Peer Interactive Communication” in scientific publications. *Journal of the American Society for Information Science and Technology*, 63(8): 1647-1661.

Costas, R., Yegros-Yegros, A. (2013). Possibilities of Funding Acknowledgement analysis for the bibliometric study of research funding organizations: case study of the Austrian Science Fund (FWF). *Proceedings of ISSI 2013 Vienna*, vol. 2. Vienna: AIT Austrian Institute of Technology.

Glänzel, W.; Thijs, B.; Debackere, K. (2013). The application of citation-based performance classes in institutional research assessment. IN: *Proceedings of 18th International Conference on Science and Technology Indicators*, Berlin, Germany. S. Hinze & A. Lottmann (Eds.), pp. 122-129.

Karlsson, S. & Persson, O. (2012). *The Swedish production of highly cited papers*. Stockholm: Vetenkapsradet: http://kva.se/Documents/Vetenskap_samhallet/Forskningspolitik/2012/akademirapport_breakthrough_research_121209.pdf.

Martin, B. & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy*, vol. 2(12), 61-90.

Moed, H.F. (2005). *Citation analysis in research evaluation*. Springer.

Ruiz-Castillo, J. (2011). The evaluation of citation distributions. Working papers. Universidad Carlos III de Madrid. http://www.eco.uc3m.es/temp/11-12.pdf.

Schneider, J. W. (2013). Caveats for using statistical significance test in research assessments. *Journal of Informetrics*, 7(1) 50-62.

Schubert, A.; Glänzel, W. and Braun, T. (1987). Subject field characteristics citation scores and scales for assessing research performance. *Scientometrics*, 12(5): 267-292.

Waltman, L., Calero-Medina, C., Noyons, E.C.M, Tijssen, R.J.W., van Eck, N.J., van Leeuwen, T.N., van Raan, A.F.J., Visser, M, and Wouters P. (2012) The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63, 12, 2419-2432.

Waltman, L. & Schreiber, M. (2012). On the calculation of percentile-based bibliometric indicators. Arxiv: http://arxiv.org/abs/1205.0646.

Waltman, L., van Eck, N.J., van Leeuwen, T.N., Visser, M.S., van Raan, A.F.J. (2011a) Towards a new crown indicator: Some theoretical considerations, *Journal of Informetrics*, 5 (1), 37-47.

Waltman, L., van Eck, N.J., van Leeuwen, T.N., Visser, M.S., van Raan, A.F.J. (2011b). Towards a new crown indicator: An empirical analysis. *Scientometrics*, 87 (3), 467-481.

Waltman, L.; Van Eck, N.J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12): 2378-2392.

# APPENDICIES

## Appendix I: NOWT-classification.

| Main_category | Subject_category |
|---|---|
| ENGINEERING SCIENCES | ACOUSTICS |
| ENGINEERING SCIENCES | AUTOMATION & CONTROL SYSTEMS |
| ENGINEERING SCIENCES | CONSTRUCTION & BUILDING TECHNOLOGY |
| ENGINEERING SCIENCES | ENERGY & FUELS |
| ENGINEERING SCIENCES | ENGINEERING, AEROSPACE |
| ENGINEERING SCIENCES | ENGINEERING, CIVIL |
| ENGINEERING SCIENCES | ENGINEERING, ELECTRICAL & ELECTRONIC |
| ENGINEERING SCIENCES | ENGINEERING, INDUSTRIAL |
| ENGINEERING SCIENCES | ENGINEERING, MANUFACTURING |
| ENGINEERING SCIENCES | ENGINEERING, MECHANICAL |
| ENGINEERING SCIENCES | ENGINEERING, MULTIDISCIPLINARY |
| ENGINEERING SCIENCES | ENGINEERING, PETROLEUM |
| ENGINEERING SCIENCES | ERGONOMICS |
| ENGINEERING SCIENCES | INSTRUMENTS & INSTRUMENTATION |
| ENGINEERING SCIENCES | MECHANICS |
| ENGINEERING SCIENCES | MICROSCOPY |
| ENGINEERING SCIENCES | MINING & MINERAL PROCESSING |
| ENGINEERING SCIENCES | NUCLEAR SCIENCE & TECHNOLOGY |
| ENGINEERING SCIENCES | ROBOTICS |
| ENGINEERING SCIENCES | TELECOMMUNICATIONS |
| ENGINEERING SCIENCES | THERMODYNAMICS |
| ENGINEERING SCIENCES | TRANSPORTATION |
| ENGINEERING SCIENCES | TRANSPORTATION SCIENCE & TECHNOLOGY |
| MEDICAL AND LIFE SCIENCES | AGRICULTURAL ENGINEERING |
| MEDICAL AND LIFE SCIENCES | AGRICULTURAL EXPERIMENT STATION REPORTS |
| MEDICAL AND LIFE SCIENCES | AGRICULTURE, DAIRY & ANIMAL SCIENCE |
| MEDICAL AND LIFE SCIENCES | AGRICULTURE, MULTIDISCIPLINARY |
| MEDICAL AND LIFE SCIENCES | AGRONOMY |
| MEDICAL AND LIFE SCIENCES | ALLERGY |
| MEDICAL AND LIFE SCIENCES | ANATOMY & MORPHOLOGY |
| MEDICAL AND LIFE SCIENCES | ANDROLOGY |
| MEDICAL AND LIFE SCIENCES | ANESTHESIOLOGY |
| MEDICAL AND LIFE SCIENCES | AUDIOLOGY & SPEECH-LANGUAGE PATHOLOGY |
| MEDICAL AND LIFE SCIENCES | BEHAVIORAL SCIENCES |
| MEDICAL AND LIFE SCIENCES | BIOCHEMICAL RESEARCH METHODS |
| MEDICAL AND LIFE SCIENCES | BIOCHEMISTRY & MOLECULAR BIOLOGY |
| MEDICAL AND LIFE SCIENCES | BIOLOGY |
| MEDICAL AND LIFE SCIENCES | BIOPHYSICS |
| MEDICAL AND LIFE SCIENCES | BIOTECHNOLOGY & APPLIED MICROBIOLOGY |

| Main_category | Subject_category |
|---|---|
| MEDICAL AND LIFE SCIENCES | CARDIAC & CARDIOVASCULAR SYSTEMS |
| MEDICAL AND LIFE SCIENCES | CELL & TISSUE ENGINEERING |
| MEDICAL AND LIFE SCIENCES | CELL BIOLOGY |
| MEDICAL AND LIFE SCIENCES | CHEMISTRY, MEDICINAL |
| MEDICAL AND LIFE SCIENCES | CLINICAL NEUROLOGY |
| MEDICAL AND LIFE SCIENCES | CRITICAL CARE MEDICINE |
| MEDICAL AND LIFE SCIENCES | DENTISTRY/ORAL SURGERY & MEDICINE |
| MEDICAL AND LIFE SCIENCES | DERMATOLOGY |
| MEDICAL AND LIFE SCIENCES | DEVELOPMENTAL BIOLOGY |
| MEDICAL AND LIFE SCIENCES | EMERGENCY MEDICINE |
| MEDICAL AND LIFE SCIENCES | ENDOCRINOLOGY & METABOLISM |
| MEDICAL AND LIFE SCIENCES | ENGINEERING, BIOMEDICAL |
| MEDICAL AND LIFE SCIENCES | ENTOMOLOGY |
| MEDICAL AND LIFE SCIENCES | EVOLUTIONARY BIOLOGY |
| MEDICAL AND LIFE SCIENCES | FISHERIES |
| MEDICAL AND LIFE SCIENCES | FOOD SCIENCE & TECHNOLOGY |
| MEDICAL AND LIFE SCIENCES | GASTROENTEROLOGY & HEPATOLOGY |
| MEDICAL AND LIFE SCIENCES | GENETICS & HEREDITY |
| MEDICAL AND LIFE SCIENCES | GERIATRICS & GERONTOLOGY |
| MEDICAL AND LIFE SCIENCES | GERONTOLOGY |
| MEDICAL AND LIFE SCIENCES | HEALTH CARE SCIENCES & SERVICES |
| MEDICAL AND LIFE SCIENCES | HEALTH POLICY & SERVICES |
| MEDICAL AND LIFE SCIENCES | HEMATOLOGY |
| MEDICAL AND LIFE SCIENCES | HORTICULTURE |
| MEDICAL AND LIFE SCIENCES | IMMUNOLOGY |
| MEDICAL AND LIFE SCIENCES | INFECTIOUS DISEASES |
| MEDICAL AND LIFE SCIENCES | INTEGRATIVE & COMPLEMENTARY MEDICINE |
| MEDICAL AND LIFE SCIENCES | MARINE & FRESHWATER BIOLOGY |
| MEDICAL AND LIFE SCIENCES | MATERIALS SCIENCE, BIOMATERIALS |
| MEDICAL AND LIFE SCIENCES | MATHEMATICAL & COMPUTATIONAL BIOLOGY |
| MEDICAL AND LIFE SCIENCES | MEDICAL INFORMATICS |
| MEDICAL AND LIFE SCIENCES | MEDICAL LABORATORY TECHNOLOGY |
| MEDICAL AND LIFE SCIENCES | MEDICINE, GENERAL & INTERNAL |
| MEDICAL AND LIFE SCIENCES | MEDICINE, RESEARCH & EXPERIMENTAL |
| MEDICAL AND LIFE SCIENCES | MICROBIOLOGY |
| MEDICAL AND LIFE SCIENCES | MYCOLOGY |
| MEDICAL AND LIFE SCIENCES | NEUROIMAGING |
| MEDICAL AND LIFE SCIENCES | NEUROSCIENCES |
| MEDICAL AND LIFE SCIENCES | NURSING |
| MEDICAL AND LIFE SCIENCES | NUTRITION & DIETETICS |
| MEDICAL AND LIFE SCIENCES | OBSTETRICS & GYNECOLOGY |
| MEDICAL AND LIFE SCIENCES | ONCOLOGY |

| Main_category | Subject_category |
|---|---|
| MEDICAL AND LIFE SCIENCES | OPHTHALMOLOGY |
| MEDICAL AND LIFE SCIENCES | ORNITHOLOGY |
| MEDICAL AND LIFE SCIENCES | ORTHOPEDICS |
| MEDICAL AND LIFE SCIENCES | OTORHINOLARYNGOLOGY |
| MEDICAL AND LIFE SCIENCES | PARASITOLOGY |
| MEDICAL AND LIFE SCIENCES | PATHOLOGY |
| MEDICAL AND LIFE SCIENCES | PEDIATRICS |
| MEDICAL AND LIFE SCIENCES | PERIPHERAL VASCULAR DISEASE |
| MEDICAL AND LIFE SCIENCES | PHARMACOLOGY & PHARMACY |
| MEDICAL AND LIFE SCIENCES | PHYSIOLOGY |
| MEDICAL AND LIFE SCIENCES | PLANT SCIENCES |
| MEDICAL AND LIFE SCIENCES | PRIMARY HEALTH CARE |
| MEDICAL AND LIFE SCIENCES | PSYCHIATRY |
| MEDICAL AND LIFE SCIENCES | PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH |
| MEDICAL AND LIFE SCIENCES | RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING |
| MEDICAL AND LIFE SCIENCES | REHABILITATION |
| MEDICAL AND LIFE SCIENCES | REPRODUCTIVE BIOLOGY |
| MEDICAL AND LIFE SCIENCES | RESPIRATORY SYSTEM |
| MEDICAL AND LIFE SCIENCES | RHEUMATOLOGY |
| MEDICAL AND LIFE SCIENCES | SOCIAL WORK |
| MEDICAL AND LIFE SCIENCES | SOIL SCIENCE |
| MEDICAL AND LIFE SCIENCES | SPORT SCIENCES |
| MEDICAL AND LIFE SCIENCES | SUBSTANCE ABUSE |
| MEDICAL AND LIFE SCIENCES | SURGERY |
| MEDICAL AND LIFE SCIENCES | TOXICOLOGY |
| MEDICAL AND LIFE SCIENCES | TRANSPLANTATION |
| MEDICAL AND LIFE SCIENCES | TROPICAL MEDICINE |
| MEDICAL AND LIFE SCIENCES | UROLOGY & NEPHROLOGY |
| MEDICAL AND LIFE SCIENCES | VETERINARY SCIENCES |
| MEDICAL AND LIFE SCIENCES | VIROLOGY |
| MEDICAL AND LIFE SCIENCES | ZOOLOGY |
| MULTIDISCIPLINARY JOURNALS | MULTIDISCIPLINARY SCIENCES |
| NATURAL SCIENCES | ASTRONOMY & ASTROPHYSICS |
| NATURAL SCIENCES | BIODIVERSITY CONSERVATION |
| NATURAL SCIENCES | CHEMISTRY, ANALYTICAL |
| NATURAL SCIENCES | CHEMISTRY, APPLIED |
| NATURAL SCIENCES | CHEMISTRY, INORGANIC & NUCLEAR |
| NATURAL SCIENCES | CHEMISTRY, MULTIDISCIPLINARY |
| NATURAL SCIENCES | CHEMISTRY, ORGANIC |
| NATURAL SCIENCES | CHEMISTRY, PHYSICAL |
| NATURAL SCIENCES | COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE |
| NATURAL SCIENCES | COMPUTER SCIENCE, CYBERNETICS |

| Main_category | Subject_category |
|---|---|
| NATURAL SCIENCES | COMPUTER SCIENCE, HARDWARE & ARCHITECTURE |
| NATURAL SCIENCES | COMPUTER SCIENCE, INFORMATION SYSTEMS |
| NATURAL SCIENCES | COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS |
| NATURAL SCIENCES | COMPUTER SCIENCE, SOFTWARE ENGINEERING |
| NATURAL SCIENCES | COMPUTER SCIENCE, THEORY & METHODS |
| NATURAL SCIENCES | CRYSTALLOGRAPHY |
| NATURAL SCIENCES | ECOLOGY |
| NATURAL SCIENCES | ELECTROCHEMISTRY |
| NATURAL SCIENCES | ENGINEERING, CHEMICAL |
| NATURAL SCIENCES | ENGINEERING, ENVIRONMENTAL |
| NATURAL SCIENCES | ENGINEERING, GEOLOGICAL |
| NATURAL SCIENCES | ENGINEERING, MARINE |
| NATURAL SCIENCES | ENGINEERING, OCEAN |
| NATURAL SCIENCES | ENVIRONMENTAL SCIENCES |
| NATURAL SCIENCES | ENVIRONMENTAL STUDIES |
| NATURAL SCIENCES | FORESTRY |
| NATURAL SCIENCES | GEOCHEMISTRY & GEOPHYSICS |
| NATURAL SCIENCES | GEOGRAPHY |
| NATURAL SCIENCES | GEOGRAPHY, PHYSICAL |
| NATURAL SCIENCES | GEOLOGY |
| NATURAL SCIENCES | GEOSCIENCES, MULTIDISCIPLINARY |
| NATURAL SCIENCES | IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY |
| NATURAL SCIENCES | LIMNOLOGY |
| NATURAL SCIENCES | LOGIC |
| NATURAL SCIENCES | MATERIALS SCIENCE, CERAMICS |
| NATURAL SCIENCES | MATERIALS SCIENCE, CHARACTERIZATION & TESTING |
| NATURAL SCIENCES | MATERIALS SCIENCE, COATINGS & FILMS |
| NATURAL SCIENCES | MATERIALS SCIENCE, COMPOSITES |
| NATURAL SCIENCES | MATERIALS SCIENCE, MULTIDISCIPLINARY |
| NATURAL SCIENCES | MATERIALS SCIENCE, PAPER & WOOD |
| NATURAL SCIENCES | MATERIALS SCIENCE, TEXTILES |
| NATURAL SCIENCES | MATHEMATICS |
| NATURAL SCIENCES | MATHEMATICS, APPLIED |
| NATURAL SCIENCES | MATHEMATICS, INTERDISCIPLINARY APPLICATIONS |
| NATURAL SCIENCES | METALLURGY & METALLURGICAL ENGINEERING |
| NATURAL SCIENCES | METEOROLOGY & ATMOSPHERIC SCIENCES |
| NATURAL SCIENCES | MINERALOGY |
| NATURAL SCIENCES | NANOSCIENCE & NANOTECHNOLOGY |
| NATURAL SCIENCES | OCEANOGRAPHY |
| NATURAL SCIENCES | OPERATIONS RESEARCH & MANAGEMENT SCIENCE |
| NATURAL SCIENCES | OPTICS |
| NATURAL SCIENCES | PALEONTOLOGY |

| Main_category | Subject_category |
|---|---|
| NATURAL SCIENCES | PHYSICS, APPLIED |
| NATURAL SCIENCES | PHYSICS, ATOMIC, MOLECULAR & CHEMICAL |
| NATURAL SCIENCES | PHYSICS, CONDENSED MATTER |
| NATURAL SCIENCES | PHYSICS, FLUIDS & PLASMAS |
| NATURAL SCIENCES | PHYSICS, MATHEMATICAL |
| NATURAL SCIENCES | PHYSICS, MULTIDISCIPLINARY |
| NATURAL SCIENCES | PHYSICS, NUCLEAR |
| NATURAL SCIENCES | PHYSICS, PARTICLES & FIELDS |
| NATURAL SCIENCES | POLYMER SCIENCE |
| NATURAL SCIENCES | REMOTE SENSING |
| NATURAL SCIENCES | SOCIAL SCIENCES, MATHEMATICAL METHODS |
| NATURAL SCIENCES | SPECTROSCOPY |
| NATURAL SCIENCES | STATISTICS & PROBABILITY |
| NATURAL SCIENCES | URBAN STUDIES |
| NATURAL SCIENCES | WATER RESOURCES |
| SOCIAL AND BEHAVIORAL SCIENCES | AGRICULTURAL ECONOMICS & POLICY |
| SOCIAL AND BEHAVIORAL SCIENCES | ANTHROPOLOGY |
| SOCIAL AND BEHAVIORAL SCIENCES | AREA STUDIES |
| SOCIAL AND BEHAVIORAL SCIENCES | BUSINESS |
| SOCIAL AND BEHAVIORAL SCIENCES | BUSINESS, FINANCE |
| SOCIAL AND BEHAVIORAL SCIENCES | CULTURAL STUDIES |
| SOCIAL AND BEHAVIORAL SCIENCES | DEMOGRAPHY |
| SOCIAL AND BEHAVIORAL SCIENCES | ECONOMICS |
| SOCIAL AND BEHAVIORAL SCIENCES | EDUCATION & EDUCATIONAL RESEARCH |
| SOCIAL AND BEHAVIORAL SCIENCES | EDUCATION, SCIENTIFIC DISCIPLINES |
| SOCIAL AND BEHAVIORAL SCIENCES | EDUCATION, SPECIAL |
| SOCIAL AND BEHAVIORAL SCIENCES | ETHNIC STUDIES |
| SOCIAL AND BEHAVIORAL SCIENCES | FAMILY STUDIES |
| SOCIAL AND BEHAVIORAL SCIENCES | HOSPITALITY, LEISURE, SPORT & TOURISM |
| SOCIAL AND BEHAVIORAL SCIENCES | INDUSTRIAL RELATIONS & LABOR |
| SOCIAL AND BEHAVIORAL SCIENCES | INTERNATIONAL RELATIONS |
| SOCIAL AND BEHAVIORAL SCIENCES | MANAGEMENT |
| SOCIAL AND BEHAVIORAL SCIENCES | PLANNING & DEVELOPMENT |
| SOCIAL AND BEHAVIORAL SCIENCES | POLITICAL SCIENCE |
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, APPLIED |
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, BIOLOGICAL |
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, CLINICAL |
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, DEVELOPMENTAL |
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, EDUCATIONAL |
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, EXPERIMENTAL |
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, MATHEMATICAL |
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, MULTIDISCIPLINARY |

| Main_category | Subject_category |
|---|---|
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, PSYCHOANALYSIS |
| SOCIAL AND BEHAVIORAL SCIENCES | PSYCHOLOGY, SOCIAL |
| SOCIAL AND BEHAVIORAL SCIENCES | PUBLIC ADMINISTRATION |
| SOCIAL AND BEHAVIORAL SCIENCES | SOCIAL ISSUES |
| SOCIAL AND BEHAVIORAL SCIENCES | SOCIAL SCIENCES, BIOMEDICAL |
| SOCIAL AND BEHAVIORAL SCIENCES | SOCIAL SCIENCES, INTERDISCIPLINARY |
| SOCIAL AND BEHAVIORAL SCIENCES | SOCIOLOGY |
| SOCIAL AND BEHAVIORAL SCIENCES | WOMEN'S STUDIES |

# Appendix II. Bibliometric scores for all units of analysis and main disciplines.

| Units | Fields | P_1997-2011 | MCS | PP (uncited) | MNCS | MNJS | PP(top 10%) | TCS | P (uncited) | TNCS | P(top 10%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DNRF analysis | ENGINEERING SCIENCES | 194 | 12,91 | 16.5% | 1,57 | 1,1 | 17,40% | 2504 | 32 | 305,3 | 34 |
| Denmark | ENGINEERING SCIENCES | 9141 | 10,55 | 17.0% | 1,58 | 1,21 | 16,60% | 96423 | 1558 | 14445,2 | 1514 |
| Denmark (No DNRF) | ENGINEERING SCIENCES | 8928 | 10,51 | 17.0% | 1,58 | 1,21 | 16,50% | 93845 | 1520 | 14118,1 | 1477 |
| HARVARD UNIV | ENGINEERING SCIENCES | 2329 | 14,8 | 14.8% | 1,93 | 1,28 | 22,30% | 34468 | 344 | 4505,8 | 520 |
| STANFORD UNIV | ENGINEERING SCIENCES | 6445 | 15,68 | 14.4% | 2,37 | 1,4 | 24,50% | 101085 | 929 | 15273,5 | 1578 |
| YALE UNIV | ENGINEERING SCIENCES | 1186 | 15,92 | 16.0% | 2 | 1,34 | 19,00% | 18879 | 190 | 2366,5 | 225 |
| UNIV CALIF SAN FRANCISCO | ENGINEERING SCIENCES | 329 | 11,39 | 14.0% | 1,48 | 1,18 | 17,00% | 3746 | 46 | 485,4 | 56 |
| MIT | ENGINEERING SCIENCES | 8465 | 13,84 | 16.5% | 2,06 | 1,34 | 21,40% | 117122 | 1394 | 17410,2 | 1814 |
| UNIV CAMBRIDGE | ENGINEERING SCIENCES | 5456 | 10,48 | 17.3% | 1,46 | 1,2 | 16,00% | 57156 | 941 | 7976,8 | 871 |
| UNIV COLL LONDON | ENGINEERING SCIENCES | 2481 | 8,32 | 19.8% | 1,28 | 1,13 | 13,80% | 20646 | 491 | 3171,8 | 343 |
| IMPERIAL COLL LONDON | ENGINEERING SCIENCES | 6350 | 8,93 | 19.4% | 1,39 | 1,18 | 14,60% | 56718 | 1231 | 8848,1 | 924 |
| UNIV LEEDS | ENGINEERING SCIENCES | 2789 | 8,32 | 19.2% | 1,29 | 1,12 | 13,60% | 23204 | 535 | 3611 | 378 |
| ECOLE POLYTECN FEDERALE L | ENGINEERING SCIENCES | 4433 | 10,93 | 18.8% | 1,66 | 1,25 | 18,50% | 48466 | 832 | 7343,1 | 821 |
| DNRF analysis | MEDICAL AND LIFE SCIENCES | 5353 | 29,8 | 4.1% | 1,52 | 1,27 | 16,20% | 159523 | 219 | 8117,3 | 867 |
| Denmark | MEDICAL AND LIFE SCIENCES | 86581 | 21,27 | 7.4% | 1,33 | 1,15 | 14,30% | 1841755 | 6395 | 114929,9 | 12378 |
| Denmark (No DNRF) | MEDICAL AND LIFE SCIENCES | 80954 | 20,75 | 7.6% | 1,32 | 1,14 | 14,20% | 1680013 | 6132 | 106459,4 | 11471 |
| HARVARD UNIV | MEDICAL AND LIFE SCIENCES | 121869 | 37,62 | 5.1% | 1,92 | 1,61 | 23,40% | 4584734 | 6175 | 233424,9 | 28500 |
| STANFORD UNIV | MEDICAL AND LIFE SCIENCES | 37469 | 34,95 | 5.3% | 1,8 | 1,51 | 21,80% | 1309680 | 1986 | 67409,6 | 8166 |
| YALE UNIV | MEDICAL AND LIFE SCIENCES | 36161 | 32,91 | 5.5% | 1,65 | 1,47 | 19,40% | 1190198 | 1980 | 59581,5 | 7013 |
| UNIV CALIF SAN FRANCISCO | MEDICAL AND LIFE SCIENCES | 51366 | 35,04 | 5.0% | 1,84 | 1,52 | 22,30% | 1799846 | 2573 | 94684,1 | 11452 |
| MIT | MEDICAL AND LIFE SCIENCES | 12696 | 48,65 | 3.5% | 2,34 | 1,78 | 28,10% | 617646 | 448 | 29715,1 | 3562 |
| UNIV CAMBRIDGE | MEDICAL AND LIFE SCIENCES | 29024 | 30,89 | 5.2% | 1,65 | 1,4 | 18,90% | 896549 | 1497 | 47916,4 | 5485 |
| UNIV COLL LONDON | MEDICAL AND LIFE SCIENCES | 46184 | 28,48 | 5.8% | 1,52 | 1,32 | 17,60% | 1315404 | 2698 | 70034,6 | 8148 |
| IMPERIAL COLL LONDON | MEDICAL AND LIFE SCIENCES | 32420 | 27,03 | 5.7% | 1,5 | 1,33 | 17,00% | 876227 | 1845 | 48601,3 | 5496 |
| UNIV LEEDS | MEDICAL AND LIFE SCIENCES | 15158 | 21,75 | 6.8% | 1,31 | 1,21 | 14,20% | 329738 | 1028 | 19815,6 | 2158 |
| ECOLE POLYTECN FEDERALE L | MEDICAL AND LIFE SCIENCES | 3408 | 23,66 | 5.6% | 1,87 | 1,38 | 19,60% | 80642 | 189 | 6378,7 | 667 |
| DNRF analysis | MULTIDISCIPLINARY JOURNAL | 340 | 98,38 | 5.3% | 7,05 | 5,88 | 60,20% | 33450 | 18 | 2396,3 | 205 |
| Denmark | MULTIDISCIPLINARY JOURNAL | 1862 | 64,25 | 9.0% | 4,86 | 4,16 | 41,50% | 119639 | 168 | 9043,3 | 773 |
| Denmark (No DNRF) | MULTIDISCIPLINARY JOURNAL | 1490 | 57,36 | 9.9% | 4,38 | 3,78 | 37,20% | 85460 | 147 | 6522,7 | 554 |
| HARVARD UNIV | MULTIDISCIPLINARY JOURNAL | 7190 | 114,36 | 3.0% | 6,14 | 5,13 | 55,60% | 822237 | 212 | 44121,2 | 3996 |
| STANFORD UNIV | MULTIDISCIPLINARY JOURNAL | 2979 | 127,29 | 2.9% | 7,14 | 5,46 | 59,30% | 379193 | 85 | 21259 | 1765 |
| YALE UNIV | MULTIDISCIPLINARY JOURNAL | 2255 | 99,83 | 3.4% | 5,16 | 4,98 | 49,10% | 225109 | 77 | 11632,7 | 1106 |
| UNIV CALIF SAN FRANCISCO | MULTIDISCIPLINARY JOURNAL | 2136 | 107,5 | 2.8% | 5,04 | 4,85 | 54,80% | 229617 | 60 | 10757,8 | 1170 |
| MIT | MULTIDISCIPLINARY JOURNAL | 2543 | 131,65 | 3.0% | 7,65 | 5,9 | 60,60% | 334795 | 75 | 19453,2 | 1541 |
| UNIV CAMBRIDGE | MULTIDISCIPLINARY JOURNAL | 2176 | 82,04 | 7.1% | 5,29 | 4,8 | 47,10% | 178514 | 155 | 11514,5 | 1024 |

| Units | Fields | P_1997-2011 | MCS | PP (uncited) | MNCS | MNJS | PP(top 10%) | TCS | P (uncited) | TNCS | P(top 10%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| UNIV COLL LONDON | MULTIDISCIPLINARY JOURNAL | 1567 | 69,93 | 6.5% | 4,03 | 4,51 | 43,00% | 109579 | 102 | 6315,1 | 673 |
| IMPERIAL COLL LONDON | MULTIDISCIPLINARY JOURNAL | 1380 | 66,69 | 8.2% | 5,31 | 4,18 | 42,60% | 92039 | 113 | 7325,7 | 588 |
| UNIV LEEDS | MULTIDISCIPLINARY JOURNAL | 418 | 66,53 | 11.7% | 4,94 | 4,21 | 41,40% | 27811 | 49 | 2065,6 | 173 |
| ECOLE POLYTECN FEDERALE L | MULTIDISCIPLINARY JOURNAL | 341 | 67,42 | 6.7% | 6,66 | 4,98 | 48,00% | 22990 | 23 | 2270,1 | 164 |
| DNRF analysis | NATURAL SCIENCES | 4708 | 20,75 | 7.9% | 1,86 | 1,46 | 21,70% | 97689 | 373 | 8750,2 | 1019 |
| Denmark | NATURAL SCIENCES | 49267 | 14,94 | 13.3% | 1,37 | 1,22 | 14,90% | 735828 | 6532 | 67363,4 | 7322 |
| Denmark (No DNRF) | NATURAL SCIENCES | 44326 | 14,32 | 13.8% | 1,31 | 1,19 | 14,10% | 634758 | 6124 | 58132,8 | 6246 |
| HARVARD UNIV | NATURAL SCIENCES | 29539 | 27,3 | 8.1% | 2,12 | 1,55 | 24,80% | 806417 | 2405 | 62563,8 | 7324 |
| STANFORD UNIV | NATURAL SCIENCES | 25961 | 23,82 | 10.6% | 2,32 | 1,54 | 25,20% | 618351 | 2760 | 60145,7 | 6540 |
| YALE UNIV | NATURAL SCIENCES | 12536 | 23,72 | 10.3% | 1,99 | 1,53 | 22,80% | 297313 | 1286 | 24936,9 | 2859 |
| UNIV CALIF SAN FRANCISCO | NATURAL SCIENCES | 1823 | 37,27 | 8.9% | 2,58 | 1,53 | 20,60% | 67943 | 162 | 4707,7 | 376 |
| MIT | NATURAL SCIENCES | 33818 | 24,23 | 10.6% | 2,21 | 1,55 | 24,90% | 819378 | 3584 | 74706,7 | 8427 |
| UNIV CAMBRIDGE | NATURAL SCIENCES | 35909 | 20,85 | 11.8% | 1,79 | 1,31 | 18,20% | 748768 | 4235 | 64265,3 | 6550 |
| UNIV COLL LONDON | NATURAL SCIENCES | 16703 | 14,08 | 12.9% | 1,33 | 1,23 | 13,70% | 235142 | 2148 | 22234,6 | 2284 |
| IMPERIAL COLL LONDON | NATURAL SCIENCES | 25570 | 17,07 | 12.8% | 1,55 | 1,27 | 17,00% | 436460 | 3260 | 39746,6 | 4353 |
| UNIV LEEDS | NATURAL SCIENCES | 13133 | 13,18 | 13.3% | 1,23 | 1,21 | 13,40% | 173097 | 1740 | 16173,6 | 1759 |
| ECOLE POLYTECN FEDERALE L | NATURAL SCIENCES | 16784 | 17,33 | 14.4% | 1,74 | 1,35 | 19,40% | 290897 | 2409 | 29245,2 | 3252 |
| DNRF analysis | SOCIAL AND BEHAVIORAL SCI | 376 | 8,7 | 12.0% | 1,24 | 1,23 | 13,80% | 3271 | 45 | 467,4 | 52 |
| Denmark | SOCIAL AND BEHAVIORAL SCI | 6638 | 8,34 | 21.5% | 1,17 | 1,09 | 12,20% | 55354 | 1430 | 7736,7 | 809 |
| Denmark (No DNRF) | SOCIAL AND BEHAVIORAL SCI | 6253 | 8,32 | 22.1% | 1,16 | 1,08 | 12,10% | 52046 | 1384 | 7259,4 | 756 |
| HARVARD UNIV | SOCIAL AND BEHAVIORAL SCI | 14079 | 23,61 | 11.6% | 2,25 | 1,66 | 25,30% | 332439 | 1626 | 31733,1 | 3558 |
| STANFORD UNIV | SOCIAL AND BEHAVIORAL SCI | 7539 | 24,77 | 10.6% | 2,28 | 1,76 | 26,40% | 186766 | 795 | 17212,9 | 1993 |
| YALE UNIV | SOCIAL AND BEHAVIORAL SCI | 6625 | 19,21 | 12.5% | 1,85 | 1,56 | 22,30% | 127248 | 826 | 12252,8 | 1476 |
| UNIV CALIF SAN FRANCISCO | SOCIAL AND BEHAVIORAL SCI | 2678 | 15,29 | 10.0% | 1,42 | 1,25 | 16,20% | 40943 | 268 | 3805,3 | 435 |
| MIT | SOCIAL AND BEHAVIORAL SCI | 3736 | 27,65 | 10.3% | 2,65 | 2 | 32,10% | 103299 | 383 | 9914 | 1198 |
| UNIV CAMBRIDGE | SOCIAL AND BEHAVIORAL SCI | 4938 | 15,86 | 15.6% | 1,51 | 1,2 | 17,30% | 78299 | 771 | 7460,7 | 853 |
| UNIV COLL LONDON | SOCIAL AND BEHAVIORAL SCI | 5789 | 18,28 | 11.7% | 1,53 | 1,3 | 19,00% | 105823 | 676 | 8865,3 | 1102 |
| IMPERIAL COLL LONDON | SOCIAL AND BEHAVIORAL SCI | 1508 | 12,86 | 13.3% | 1,43 | 1,23 | 16,00% | 19392 | 201 | 2158,9 | 242 |
| UNIV LEEDS | SOCIAL AND BEHAVIORAL SCI | 2720 | 10,59 | 15.9% | 1,23 | 1,11 | 13,40% | 28805 | 433 | 3354 | 365 |
| ECOLE POLYTECN FEDERALE L | SOCIAL AND BEHAVIORAL SCI | 239 | 8,89 | 16.3% | 1,52 | 1,42 | 18,50% | 2125 | 39 | 363,2 | 44 |

## Appendix III. Main results by percentile classes for all the units of analysis.

| Unit | percentile | P | P accumulative in percentile | P exclusive in percentile | % absolute | % exclusive |
|---|---|---|---|---|---|---|
| DNRF (analysis) | Percentile99 | 10342 | 312,01 | 312,01 | 3% | 3% |
| DNRF (analysis) | Percentile95 | 10342 | 1195,83 | 883,81 | 12% | 9% |
| DNRF (analysis) | Percentile90 | 10342 | 2101,30 | 905,47 | 20% | 9% |
| DNRF (analysis) | Percentile75 | 10342 | 4334,38 | 2233,08 | 42% | 22% |
| DNRF (analysis) | Percentile50 | 10342 | 7219,07 | 2884,70 | 70% | 28% |
| DNRF (analysis) | Percentile less 50 | 10342 | 10342 | 3122,93 | 100% | 30% |
| Denmark | Percentile99 | 140871 | 2515,49 | 2515,49 | 2% | 2% |
| Denmark | Percentile95 | 140871 | 11008,87 | 8493,38 | 8% | 6% |
| Denmark | Percentile90 | 140871 | 20929,54 | 9920,67 | 15% | 7% |
| Denmark | Percentile75 | 140871 | 48021,28 | 27091,74 | 34% | 19% |
| Denmark | Percentile50 | 140871 | 86819,67 | 38798,39 | 62% | 28% |
| Denmark | Percentile less 50 | 140871 | 140871 | 54051,33 | 100% | 38% |
| Denmark (No DNRF) | Percentile99 | 130012 | 2190,93 | 2190,93 | 2% | 2% |
| Denmark (No DNRF) | Percentile95 | 130012 | 9752,20 | 7561,27 | 8% | 6% |
| Denmark (No DNRF) | Percentile90 | 130012 | 18718,48 | 8966,28 | 14% | 7% |
| Denmark (No DNRF) | Percentile75 | 130012 | 43462,28 | 24743,80 | 33% | 19% |
| Denmark (No DNRF) | Percentile50 | 130012 | 79234,27 | 35771,99 | 61% | 28% |
| Denmark (No DNRF) | Percentile less 50 | 130012 | 130012 | 50777,73 | 100% | 39% |
| ECOLE POLYTECN FEDERALE LAUSANNE | Percentile99 | 21777 | 600,69 | 600,69 | 3% | 3% |
| ECOLE POLYTECN FEDERALE LAUSANNE | Percentile95 | 21777 | 2385,74 | 1785,06 | 11% | 8% |
| ECOLE POLYTECN FEDERALE LAUSANNE | Percentile90 | 21777 | 4290,37 | 1904,62 | 20% | 9% |
| ECOLE POLYTECN FEDERALE LAUSANNE | Percentile75 | 21777 | 8889,76 | 4599,39 | 41% | 21% |
| ECOLE POLYTECN FEDERALE LAUSANNE | Percentile50 | 21777 | 14709,25 | 5819,49 | 68% | 27% |
| ECOLE POLYTECN FEDERALE LAUSANNE | Percentile less 50 | 21777 | 21777 | 7067,75 | 100% | 32% |
| HARVARD UNIV | Percentile99 | 169013 | 7074,90 | 7074,90 | 4% | 4% |
| HARVARD UNIV | Percentile95 | 169013 | 25358,57 | 18283,67 | 15% | 11% |
| HARVARD UNIV | Percentile90 | 169013 | 42529,29 | 17170,73 | 25% | 10% |
| HARVARD UNIV | Percentile75 | 169013 | 79956,75 | 37427,46 | 47% | 22% |
| HARVARD UNIV | Percentile50 | 169013 | 122024,05 | 42067,29 | 72% | 25% |
| HARVARD UNIV | Percentile less 50 | 169013 | 169013 | 46988,95 | 100% | 28% |
| IMPERIAL COLL LONDON | Percentile99 | 61223 | 1401,37 | 1401,37 | 2% | 2% |
| IMPERIAL COLL LONDON | Percentile95 | 61223 | 5823,04 | 4421,67 | 10% | 7% |
| IMPERIAL COLL LONDON | Percentile90 | 61223 | 10620,68 | 4797,64 | 17% | 8% |
| IMPERIAL COLL LONDON | Percentile75 | 61223 | 23379,53 | 12758,85 | 38% | 21% |
| IMPERIAL COLL LONDON | Percentile50 | 61223 | 40075,71 | 16696,18 | 65% | 27% |
| IMPERIAL COLL LONDON | Percentile less 50 | 61223 | 61223 | 21147,29 | 100% | 35% |
| MIT | Percentile99 | 54939 | 2835,10 | 2835,10 | 5% | 5% |
| MIT | Percentile95 | 54939 | 9406,92 | 6571,82 | 17% | 12% |

| | | | | | | |
|---|---|---|---|---|---|---|
| MIT | Percentile90 | 54939 | 15226,81 | 5819,89 | 28% | 11% |
| MIT | Percentile75 | 54939 | 27655,00 | 12428,19 | 50% | 23% |
| MIT | Percentile50 | 54939 | 40937,90 | 13282,90 | 75% | 24% |
| MIT | Percentile less 50 | 54939 | 54939 | 14001,10 | 100% | 25% |
| STANFORD UNIV | Percentile99 | 75152 | 3240,33 | 3240,33 | 4% | 4% |
| STANFORD UNIV | Percentile95 | 75152 | 11450,65 | 8210,32 | 15% | 11% |
| STANFORD UNIV | Percentile90 | 75152 | 18976,72 | 7526,07 | 25% | 10% |
| STANFORD UNIV | Percentile75 | 75152 | 35727,07 | 16750,36 | 48% | 22% |
| STANFORD UNIV | Percentile50 | 75152 | 54416,95 | 18689,88 | 72% | 25% |
| STANFORD UNIV | Percentile less 50 | 75152 | 75152 | 20735,05 | 100% | 28% |
| UNIV CALIF SAN FRANCISCO | Percentile99 | 55735 | 1907,39 | 1907,39 | 3% | 3% |
| UNIV CALIF SAN FRANCISCO | Percentile95 | 55735 | 7446,15 | 5538,76 | 13% | 10% |
| UNIV CALIF SAN FRANCISCO | Percentile90 | 55735 | 13051,05 | 5604,91 | 23% | 10% |
| UNIV CALIF SAN FRANCISCO | Percentile75 | 55735 | 25480,28 | 12429,22 | 46% | 22% |
| UNIV CALIF SAN FRANCISCO | Percentile50 | 55735 | 39851,69 | 14371,41 | 72% | 26% |
| UNIV CALIF SAN FRANCISCO | Percentile less 50 | 55735 | 55735 | 15883,31 | 100% | 28% |
| UNIV CAMBRIDGE | Percentile99 | 74207 | 1983,14 | 1983,14 | 3% | 3% |
| UNIV CAMBRIDGE | Percentile95 | 74207 | 7858,03 | 5874,89 | 11% | 8% |
| UNIV CAMBRIDGE | Percentile90 | 74207 | 14153,39 | 6295,36 | 19% | 8% |
| UNIV CAMBRIDGE | Percentile75 | 74207 | 30116,47 | 15963,08 | 41% | 22% |
| UNIV CAMBRIDGE | Percentile50 | 74207 | 50001,67 | 19885,20 | 67% | 27% |
| UNIV CAMBRIDGE | Percentile less 50 | 74207 | 74207 | 24205,33 | 100% | 33% |
| UNIV COLL LONDON | Percentile99 | 69212 | 1420,94 | 1420,94 | 2% | 2% |
| UNIV COLL LONDON | Percentile95 | 69212 | 6378,86 | 4957,92 | 9% | 7% |
| UNIV COLL LONDON | Percentile90 | 69212 | 11920,50 | 5541,64 | 17% | 8% |
| UNIV COLL LONDON | Percentile75 | 69212 | 26251,45 | 14330,95 | 38% | 21% |
| UNIV COLL LONDON | Percentile50 | 69212 | 45060,87 | 18809,42 | 65% | 27% |
| UNIV COLL LONDON | Percentile less 50 | 69212 | 69212 | 24151,13 | 100% | 35% |
| UNIV LEEDS | Percentile99 | 32061 | 472,79 | 472,79 | 1% | 1% |
| UNIV LEEDS | Percentile95 | 32061 | 2293,13 | 1820,35 | 7% | 6% |
| UNIV LEEDS | Percentile90 | 32061 | 4499,09 | 2205,96 | 14% | 7% |
| UNIV LEEDS | Percentile75 | 32061 | 10725,98 | 6226,88 | 33% | 19% |
| UNIV LEEDS | Percentile50 | 32061 | 19780,70 | 9054,72 | 62% | 28% |
| UNIV LEEDS | Percentile less 50 | 32061 | 32061 | 12280,30 | 100% | 38% |
| YALE UNIV | Percentile99 | 57071 | 1809,02 | 1809,02 | 3% | 3% |
| YALE UNIV | Percentile95 | 57071 | 7085,06 | 5276,04 | 12% | 9% |
| YALE UNIV | Percentile90 | 57071 | 12445,51 | 5360,46 | 22% | 9% |
| YALE UNIV | Percentile75 | 57071 | 24867,19 | 12421,68 | 44% | 22% |
| YALE UNIV | Percentile50 | 57071 | 39633,99 | 14766,80 | 69% | 26% |
| YALE UNIV | Percentile less 50 | 57071 | 57071 | 17437,01 | 100% | 31% |

## Appendix IV. High impact journal publications – main results.

| unit | Totl P | P in HIJ | % in HIJ | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DNRF (analysis) | 10342 | 1351 | 13% | 47 | 53 | 57 | 69 | 85 | 69 | 74 | 70 | 92 | 98 | 110 | 115 | 121 | 149 | 142 |
| Denmark | 140871 | 10636 | 8% | 478 | 515 | 574 | 526 | 584 | 582 | 586 | 627 | 612 | 672 | 814 | 941 | 987 | 1001 | 1137 |
| Denmark (No DNRF) | 130012 | 9204 | 7% | 431 | 462 | 517 | 457 | 499 | 513 | 512 | 557 | 520 | 574 | 704 | 815 | 845 | 827 | 971 |
| ECOLE POLYTECN FEDERALE LAUSANNE | 21777 | 2497 | 11% | 67 | 71 | 82 | 84 | 74 | 98 | 101 | 166 | 163 | 217 | 224 | 263 | 258 | 303 | 326 |
| HARVARD UNIV | 169013 | 36109 | 21% | 1788 | 1835 | 1853 | 1913 | 2058 | 2023 | 1937 | 2199 | 2319 | 2544 | 2858 | 3038 | 3174 | 3286 | 3284 |
| IMPERIAL COLL LONDON | 61223 | 6895 | 11% | 208 | 257 | 265 | 271 | 332 | 354 | 349 | 420 | 459 | 468 | 582 | 710 | 687 | 763 | 770 |
| MIT | 54939 | 11628 | 21% | 621 | 606 | 632 | 593 | 619 | 670 | 652 | 741 | 767 | 827 | 940 | 900 | 985 | 1001 | 1074 |
| STANFORD UNIV | 75152 | 14799 | 20% | 759 | 761 | 806 | 883 | 876 | 912 | 930 | 958 | 956 | 943 | 1177 | 1147 | 1145 | 1225 | 1321 |
| UNIV CALIF SAN FRANCISCO | 55735 | 10411 | 19% | 543 | 569 | 547 | 554 | 583 | 571 | 581 | 583 | 647 | 700 | 797 | 827 | 929 | 959 | 1021 |
| UNIV CAMBRIDGE | 74207 | 8729 | 12% | 377 | 432 | 403 | 442 | 442 | 463 | 497 | 558 | 572 | 571 | 665 | 767 | 804 | 817 | 919 |
| UNIV COLL LONDON | 69212 | 8243 | 12% | 312 | 322 | 368 | 375 | 487 | 468 | 438 | 442 | 532 | 549 | 694 | 733 | 757 | 861 | 905 |
| UNIV LEEDS | 32061 | 2463 | 8% | 95 | 87 | 131 | 117 | 125 | 142 | 134 | 172 | 135 | 160 | 193 | 216 | 225 | 268 | 263 |
| YALE UNIV | 57071 | 11208 | 20% | 638 | 643 | 577 | 643 | 591 | 655 | 679 | 764 | 756 | 800 | 855 | 862 | 873 | 935 | 937 |

# Appendix V. Main results for analysis of 'recruitment' rates of new scientists publishing highly cited publications within 3 years of their first WoS publication.

| Country | Successful[1] | Total | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008[2] | 2009[2] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| all countries | No | 1414056 | 61419 | 62791 | 69674 | 69140 | 72430 | 75199 | 77512 | 78808 | 80389 | 89729 | 91922 | 100548 | 97081 | 104076 | 106022 | 101538 | 75778 |
| all countries | Yes | 714018 | 30795 | 31307 | 33497 | 32588 | 33387 | 34109 | 35350 | 35441 | 37519 | 41310 | 44258 | 48284 | 48760 | 53303 | 55837 | 61053 | 57220 |
| all countries | Total | 2128074 | 92214 | 94098 | 103171 | 101728 | 105817 | 109308 | 112862 | 114249 | 117908 | 131039 | 136180 | 148832 | 145841 | 157379 | 161859 | 162591 | 132998 |
| DENMARK | No | 7709 | 359 | 406 | 408 | 376 | 442 | 439 | 426 | 446 | 428 | 488 | 495 | 544 | 501 | 527 | 550 | 502 | 372 |
| DENMARK | Yes | 5489 | 236 | 231 | 282 | 239 | 298 | 273 | 259 | 307 | 299 | 285 | 374 | 331 | 365 | 394 | 406 | 466 | 444 |
| DENMARK | Total | 13198 | 595 | 637 | 690 | 615 | 740 | 712 | 685 | 753 | 727 | 773 | 869 | 875 | 866 | 921 | 956 | 968 | 816 |
| DNRF | No | 931 | 35 | 40 | 36 | 38 | 41 | 47 | 67 | 82 | 75 | 71 | 74 | 65 | 73 | 70 | 64 | 29 | 24 |
| DNRF | Yes | 954 | 30 | 41 | 33 | 37 | 40 | 47 | 54 | 68 | 60 | 50 | 58 | 66 | 78 | 80 | 66 | 80 | 66 |
| DNRF | Total | 1885 | 65 | 81 | 69 | 75 | 81 | 94 | 121 | 150 | 135 | 121 | 132 | 131 | 151 | 150 | 130 | 109 | 90 |
| FINLAND | No | 9289 | 406 | 451 | 504 | 520 | 568 | 614 | 594 | 587 | 537 | 625 | 660 | 615 | 582 | 606 | 569 | 508 | 343 |
| FINLAND | Yes | 5210 | 239 | 273 | 300 | 324 | 298 | 258 | 338 | 283 | 272 | 339 | 298 | 327 | 328 | 340 | 334 | 371 | 288 |
| FINLAND | Total | 14499 | 645 | 724 | 804 | 844 | 866 | 872 | 932 | 870 | 809 | 964 | 958 | 942 | 910 | 946 | 903 | 879 | 631 |
| NETHERLANDS | No | 22647 | 1229 | 1180 | 1249 | 1169 | 1240 | 1267 | 1219 | 1344 | 1199 | 1439 | 1456 | 1604 | 1490 | 1598 | 1541 | 1349 | 1074 |
| NETHERLANDS | Yes | 16524 | 841 | 851 | 816 | 790 | 785 | 816 | 793 | 755 | 827 | 849 | 964 | 1066 | 1137 | 1214 | 1240 | 1373 | 1407 |
| NETHERLANDS | Total | 39171 | 2070 | 2031 | 2065 | 1959 | 2025 | 2083 | 2012 | 2099 | 2026 | 2288 | 2420 | 2670 | 2627 | 2812 | 2781 | 2722 | 2481 |
| SWEDEN | No | 16651 | 843 | 852 | 915 | 912 | 971 | 1059 | 982 | 1017 | 1113 | 1076 | 1111 | 1199 | 1129 | 1063 | 985 | 837 | 587 |
| SWEDEN | Yes | 10171 | 492 | 560 | 525 | 596 | 591 | 570 | 585 | 571 | 560 | 613 | 613 | 644 | 653 | 683 | 625 | 663 | 627 |
| SWEDEN | Total | 26822 | 1335 | 1412 | 1440 | 1508 | 1562 | 1629 | 1567 | 1588 | 1673 | 1689 | 1724 | 1843 | 1782 | 1746 | 1610 | 1500 | 1214 |
| SWITZERLAND | No | 13691 | 662 | 646 | 668 | 660 | 682 | 697 | 716 | 781 | 778 | 880 | 929 | 1025 | 1028 | 977 | 1010 | 845 | 707 |
| SWITZERLAND | Yes | 12427 | 481 | 506 | 605 | 532 | 620 | 594 | 642 | 610 | 627 | 708 | 753 | 893 | 864 | 944 | 986 | 1124 | 938 |
| SWITZERLAND | Total | 26118 | 1143 | 1152 | 1273 | 1192 | 1302 | 1291 | 1358 | 1391 | 1405 | 1588 | 1682 | 1918 | 1892 | 1921 | 1996 | 1969 | 1645 |

[1] 'Successful' – total is the number of new scientists that has his or her first publication in a particular year; yes is the number of these new scientists that also has at least one top 10% publication with the next three years (first publication year plus 2 more years).

[2] Notice, we do not show 'recruitment' rates for 2008 and 2009 in Figure 10.1 for interpretative purposes. The steep rise for all units is an artifact due to a combination of weighting tied publications around the 90$^{th}$ percentile and low citation density in the latest years investigated creating many ties with low weights that eventually will contribute to the top 10%

## Appendix VI. CSS average and standard deviation values for the 784 CWTS meso-fields

| | Type1 | Type2 | Type3 | Type4 | Citations_t1 | Citations_t2 | Citations_t3 | Citations_t4 |
|---|---|---|---|---|---|---|---|---|
| Avg. | 73.9% | 18.9% | 5.2% | 2.0% | 21.6% | 32.1% | 21.8% | 24.5% |
| St. Dev | 2.6% | 1.5% | 0.8% | 0.5% | 1.8% | 1.1% | 0.9% | 1.9% |

As it can be seen, with a relatively low standard deviation, we can see how in general type 1 publications (74% of all the publications in every meso-field) have received around 22% of all citations given in the meso-field. Type 2 publications represent 19% of all the publications in the meso-field and receive around 32% of all the citations in the field. Type 3 publications amount to around 5% of the publications of every meso-field receiving 22% of all the citations; and finally Type 4 are 2% of the most cited publications in every meso-field and they alone receive around 25% of all the citations received in the field.