

The 'recruitment' rate of new scientists who publishes highly cited publications

Jesper W. Schneider, jws@cfa.au.dk

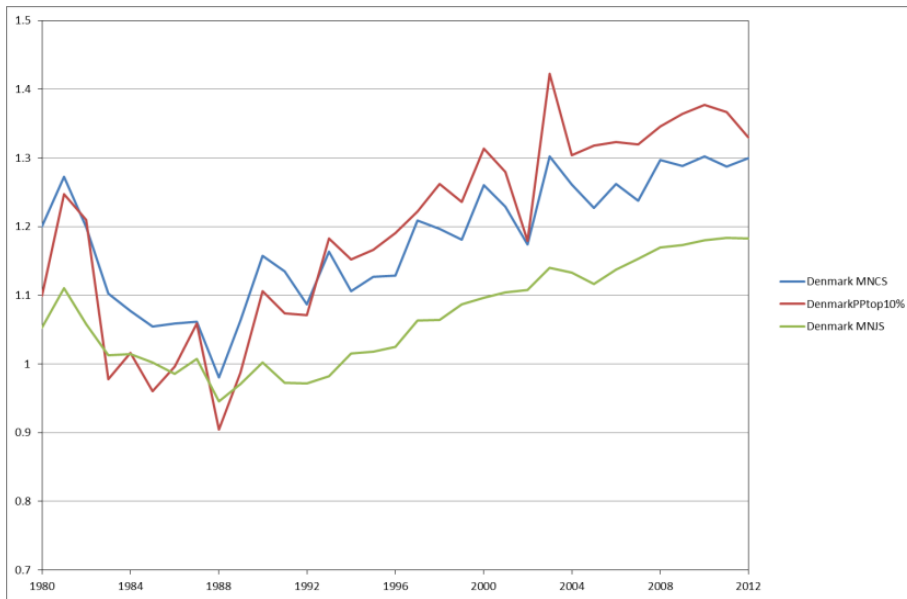
Danish Centre for Studies in Research and Research Policy, Department of Political Science and Government, Aarhus University, Denmark

Rodrigo Costas, rcostas@cwts.leidenuniv.nl

Centre for Science and Technology Studies, Leiden University, Netherlands

Background

“The Danish Wonder”



- Policy discussion
- Numerous bibliometric analyses

Background

12. RECRUITMENT OF TOP SCIENTISTS

For a country to keep pace with the outside world, it is important to be able to recruit top scientists. If not done sufficiently, we can eventually expect stagnation both in terms of research volume and attractiveness. This question can be illuminated using bibliometric data.

Let us define the scientific elite in a country as the authors in a given field who have written at least five top 10 *N*-publications over a fifteen year period. We want to identify the elite author community in each country for three, overlapping 15 year periods: 1986-2000, 1991-2005 and 1996-2010, respectively. Since the database lack a link between author names and addresses before 2006, when there are several countries among the author addresses there is some uncertainty of the nationality of an author. In order to somewhat reduce the number of elite authors with incorrect country affiliation, we restrict our search to the highly cited publications (articles and reviews) where all author addresses are from the country in focus and the publications with up to at most 3 different countries among the author addresses and with at most 10 authors. For each country, each field and each of the three 15 year periods, we identify the year the author wrote his or her first highly cited article in this field. Then we calculate the percentage of authors who made his or her debut during the last five years of the fifteen-year period in question. By this definition it is more difficult to enter the elite group for those starting to publish late in a period as compared to those starting to publish early in the period. This effect is, however, similar for all compared countries. The recruitment statistics was calculated for each SPRL subject field and each of the 250 journal fields. Both these subject groupings resulted in very similar results and only the results based on the SPRL-fields are presented. The calculations are based on the whole counts production of papers (no fractionalization is applied).

These estimates are associated with similar potential errors as mentioned in section 11, and with the restriction to articles with at most 3 countries and at most 10 authors. Again, we assume that errors are similar in all six countries compared. Thus, the recruitment rates presented here are intended to indicate relative differences among the compared countries and should not be interpreted as absolute measures of recruitment rate. Further, since the estimated recruitment-rate can be related to changes in the total output of publications and the growth of the entire author community; the mean annual growth rate of both these are presented in table 12.1.

For all countries, the recruitment rate decreases from period 1 to period 2. Between the two last periods the pattern is not consistent among countries; for four of the countries the rate continue to decline while for the other two, Switzerland and the United Kingdom, the recruitment rate increases the last period (only marginally so for the United Kingdom).

Sweden shows the lowest recruitment rate at the first period and the second lowest rate for the two others. In parallel with relatively low recruitment rates, Sweden also has the second lowest growth rate among the compared countries for the total publication volume and for the size of the entire author community, respectively.

“For a country to keep pace with the outside world, it is important to be able to recruit top scientists. If not done sufficiently, we can eventually expect stagnation both in terms of research volume and attractiveness. This question can be illuminated using bibliometric data.”

(Karlsson & Persson, 2012)

Background

“There is only one proved method of assisting the advancement of pure science – that is picking men of genius, backing them heavily, and leaving them to direct themselves.”

(President of Harvard University James Bryan Conant , Letter to the New York Times, 13. August 1945)

- Very interesting notion from a science policy, scientometric, sociology of science and psychology of science perspective
- Recruitment of young talented researchers is no doubt an important factor when it comes to producing high quality research (and perhaps raising or sustaining high impact)
- Knowing something about “recruitment rates” is certainly interesting
- But also very difficult to operationalize, both retrospectively and for prediction, especially within the restricted universe of citation databases

Background – The Swedish approach

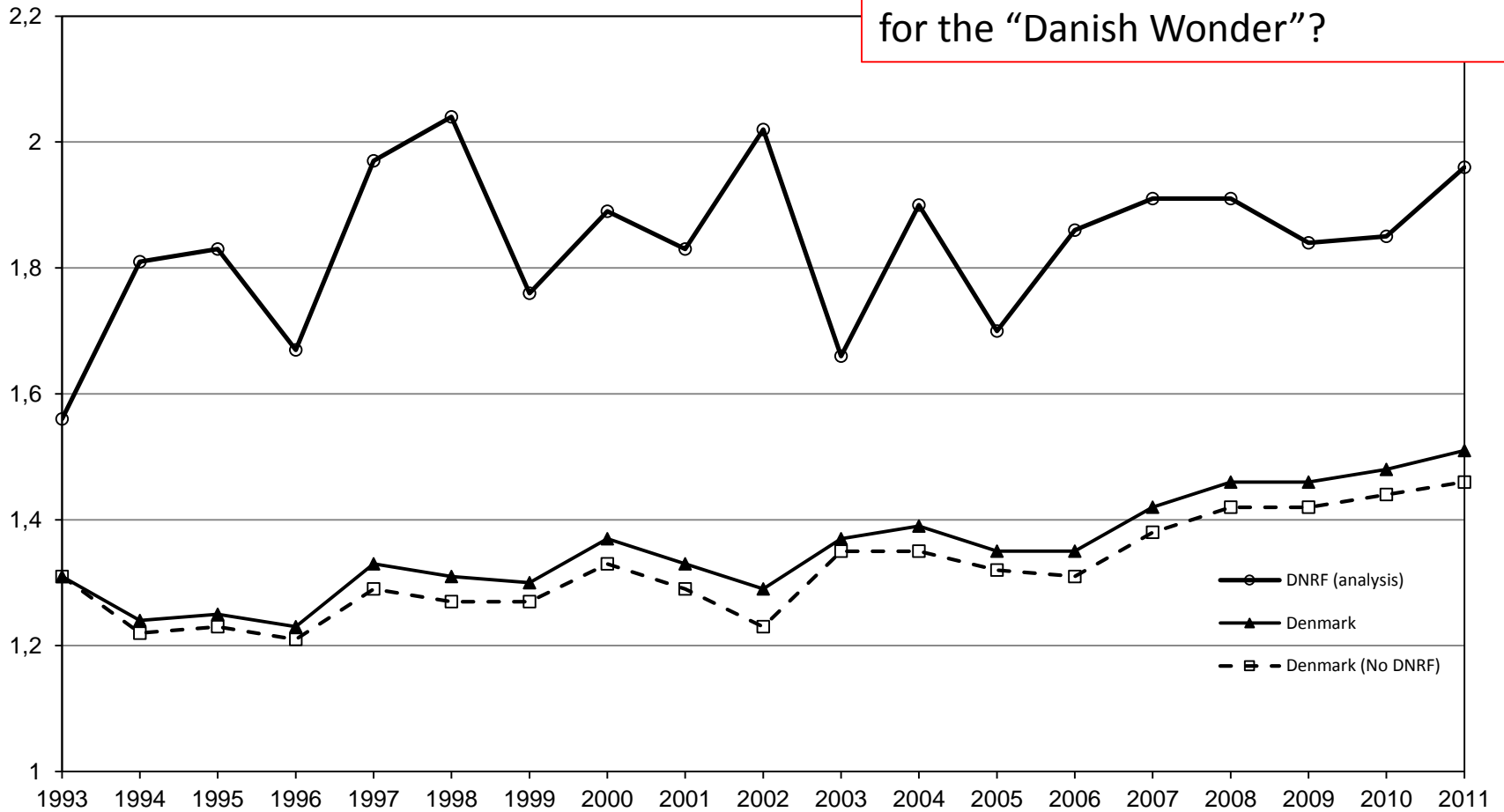
- Country level comparison of ‘recruitment rates’
- ‘Elite researchers’ = 5 top 10% publications over 15 years
- Identify ‘elite’ community for each country for three overlapping periods (1986-2000; 1991-2005; 1996-2010)
- Challenge: lack of links between author names and addresses before 2008 in WoS
- Solution = reducing the data set considerably, assuming that systematic biases are evenly distributed
 - “we restrict our search to the highly cited publications (articles and reviews) where all author addresses are from the country in focus and the publications with up to at most 3 different countries among the author addresses and with at most 10 authors”
- Calculation of ‘recruitment rate’ = identify when an author had her/his first top 10 % paper and the calculate the percentage of authors who had their debut in last five years of a 15-year period

Our approach & context

- We basically explore the same question – how to estimate the ‘recruitment rate’?
- Our context and unit of analysis is different
 - Centres of Excellence (CoE) in Denmark funded by the Danish National Research Foundation (DNRF)
 - National benchmarks
- Our approach is also very different and based on an advanced name disambiguation algorithm (Caron & van Eck, 2014)

Context: Performance of DNRF

Very high performance – a driver for the “Danish Wonder”?



Our approach - definitions

- Researchers identified = name disambiguation (Caron & van Eck, 2014)
- Top publication = top 10 % cited in a field
- 'Top' researcher = one top 10 % paper within three years of first publication
- First publication year = first publication year identified in WoS
- Author-affiliation certain linkages = 'certain' linkages between author and country
 - First author – first country
 - All authors when there is only one country
 - Reprint author and reprint country
 - Direct linkages from 2008
 - Most common country (MCC) of a researcher

Our approach – methodological steps

- Basic premise – very simple:
 - 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

Our approach – methodological steps

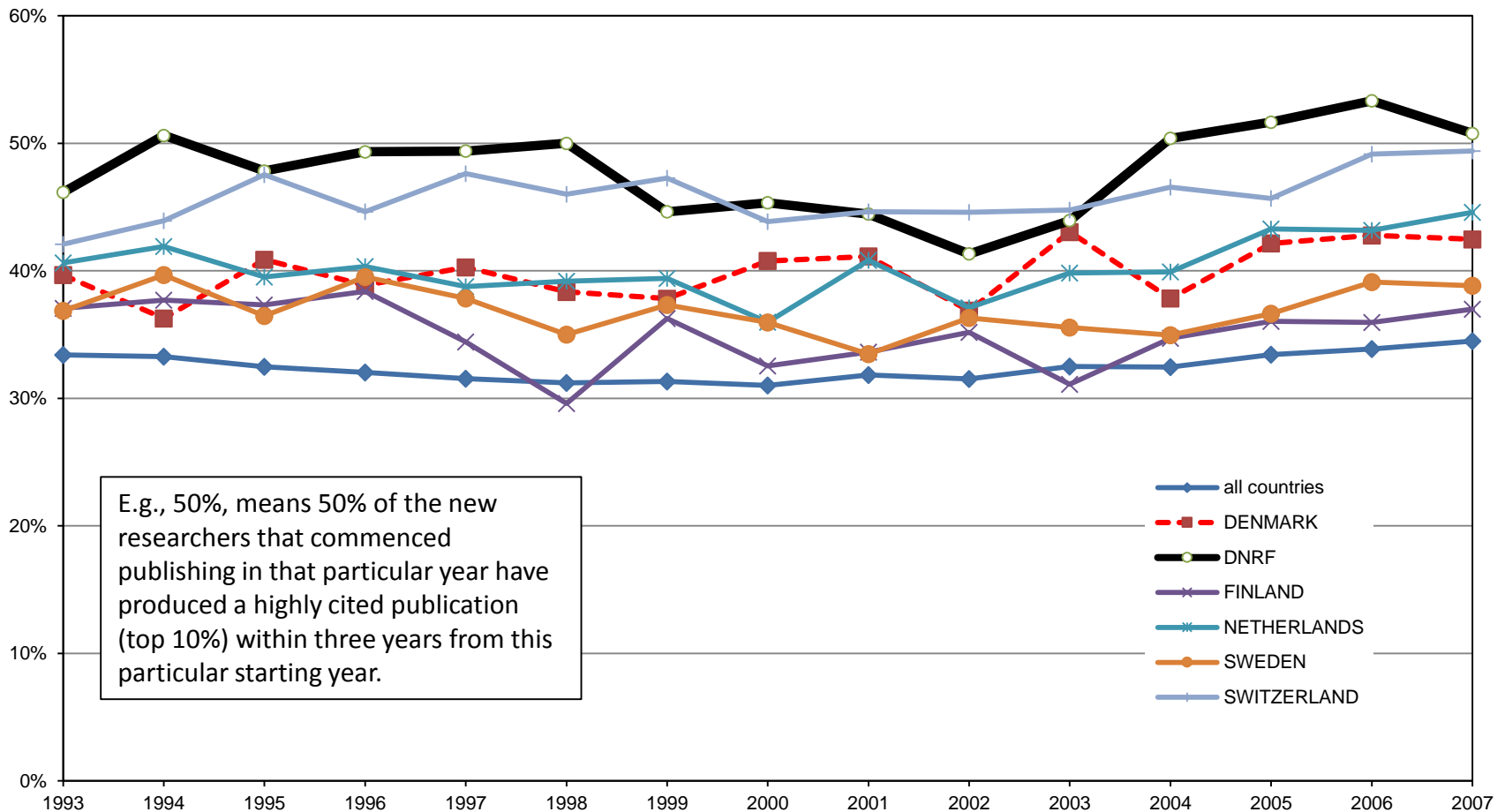
1. Selection of all publications and authors in WoS (1993-2011, Science Citation Index and the Social Sciences Citation Index) = 90.885,730 author-publication combinations and 22.871,295 unique publications
2. Detection of all the researchers active in the previous dataset and their first publication year (author-name disambiguation algorithm) = 22.377,560 unique researchers
 - We also identified their first year of publication in WoS (full oeuvre of the identified researchers, 1980-2012) – excluding authors that started to publish before the year 1993 (homogenous cohorts of researchers)
 - Exclusion of researchers with less than 5 publications in the period and those that started publishing after 2009
 - 2.128,074 researchers from all over the world meet these criteria and were included for the final analysis

Our approach – methodological steps

3. Calculation of the MCC for all identified researchers
 - 2,072,696 (97%) = at least one ‘certain’ country or MCC
4. Identification of ‘successful’ new researchers (1 top10% paper within 3 years of first publication year in WoS)
 - 714,152 (34%) ‘successful’ researchers (whole database, 1993-2009)
5. Identification of DNRF researchers: 1) MCC as ‘Denmark’ and 2) more than 10% of publications belonging to a specific set of DNRF-publications
 - 13,198 researchers (0.6 % of the total) = ‘Denmark’ as (one of) their MCC and 1,885 (14 % of Danish researcher) can be considered a ‘DNRF-researchers’
6. Estimating the ‘recruitment rate’ of new ‘successful’ researchers for Denmark, the DNRF, and a selection of benchmark countries and depicting the developments relative to the database growth

Results

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.



Reflections

- Results are as expected for countries and matches those of the Swedish report, even though we apply a very different methodology
- No reason to believe that the overall patterns will change if we changed the thresholds for highly cited papers or the number of highly cited papers needed for a researcher to be considered a 'top scientists'
- 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 performed with the same conditions for all publications and their authors in the whole database with an advanced author detection algorithm, making it highly robust

Reflections

- 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
- The methodology also has a number of limitations
 - The data quality is certainly not optimal and the author-identification algorithm is not perfect, although
 - Misclassification
 - precision and recall values of 95% and 90% respectively
 - Validity
 - Retrospective ‘recruitment rates’: We infer that high-impact performance is correlated with higher ‘recruitment rates’ ... or is it vice versa ... or is it spurious?

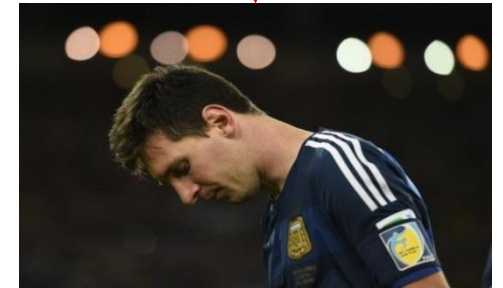
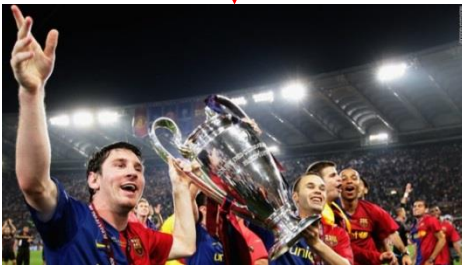
Reflections

- Actual ‘recruitment rates’ should be interpreted carefully, in as much as they to a large extent reflect collaboration practices
- 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
- Most of the papers from ‘top’ researchers have multiple authorships
 - What is the ‘top’ researchers’ contribution?
 - Are they ‘recruited’ by an already high-performing unit and thus benefit from that, or are they (their research) actually causing the paper to be highly cited and benefiting the unit?

What is more determinant, the individual or the team?



Would Messi have become so famous/successful in a different team? And would Barcelona have won so many tournaments without Messi?



Thank you for your attention!