Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht Scientometrics, Vol. 77, No. 2 (2008) 207–222 DOI: 10.1007/s11192-007-1962-y

National research contributions: A case study on Finnish biomedical research

PENTTI RIIKONEN,^a MAUNO VIHINEN^{b,c}

^a Department of Information Science, University of Turku, Turku (Finland) ^b Institute of Medical Technology, University of Tampere, Tampere (Finland) ^c Tampere University Hospital, Tampere (Finland)

The long-term influence and contribution of research can be evaluated relatively reliably by bibliometric citation analysis. Previously, productivity of nations has been estimated by using either the number of published articles or journal impact factors and/or citation data. These studies show certain trends, but detailed analysis is not possible due to the assumption that all articles in a journal were equally cited. Here we describe the first comprehensive, longterm, nationwide analysis of scientific performance. We studied the lifetime research output of 748 Finnish principal investigators in biomedicine during the years 1966–2000, analysed national trends, and made a comparison with international research production. Our results indicate that analyses of the scientific contribution of persons, disciplines, or nations should be based on actual publication and citation counts rather than on derived information like impact factors. 51% of the principal investigators have published altogether 75% of the articles; however, the whole scientific community has contributed to the growth of biomedical research in Finland since the Second World War.

Introduction

The influence and contribution of research over time can be evaluated relatively reliably by bibliometric citation analysis. In addition to their primary function, i.e., indicating the influence of an article, citations have many functions. Citations can also be persuasive, negative, or self-citations. They can also contain cultural and language

Address for correspondence: MAUNO VIHINEN Institute of Medical Technology, FI-33014 University of Tampere, Tampere, Finland E-mail: mauno.vihinen@uta.fi

0138–9130/US \$ 20.00 Copyright © 2008 Akadémiai Kiadó, Budapest All rights reserved

Received October 3, 2007

biases and reflect the availability of articles – freely accessible articles can be more visible. However, we can assume that large datasets will provide a balanced and reliable view. Here we have performed, to our knowledge, the first comprehensive, long term, nationwide analysis of scientific performance.

Many previous studies have considered national contributions based on either the journal impact factors and/or citation data for only a limited area and just for a few years [BOLDT & AL., 1999; FAVALORO, 1998; GARCIA-RIO & AL., 2001; LUUKKONEN, 1990; MELA & CIMMINO, 1998; MELA & AL., 1999; PERSSON & AL., 2000; SPERTI & IERARDI, 1999], or the number of published articles [GAGNON & AL., 2000]. Many of these studies also contain information regarding Finland. Some reports compare overall national and international scientific production [HUSSO & AL., 2000; NIKKARI, 1995; PERSSON & AL., 2000]. Although these studies show certain trends, detailed analysis is not possible due to the assumption that all articles in a journal are equally cited. Governments, industry, research institutes, funding bodies and others are interested in national output and outcomes to evaluate the productivity and quality of research.

We studied the research output and citations of 748 Finnish principal investigators (PIs) in biomedicine during the years 1966–2000. The analysed PIs are in charge of practically the whole scientific production in the country. Biomedical research is conducted in groups and reported in multiauthored articles. Since junior scientists write their publications together with their supervisors and group leaders, we just needed to collect the articles of the PIs to have an (almost) complete list of Finnish publications. In addition to an analysis of national trends, we also made a comparison with international research production. We revealed the impact of Finnish biomedical science based on the lifetime publication and citation information of all the major scientists in the field. Out of any previous investigations, the one closest to our study comes from an approach in clinical cardiology where citation frequencies were averaged over 5 years [DE JONG & SCHAPER, 1996]. This study was limited to a period of 12 years.

Finland is well suited for this kind of analysis for a number of reasons. Being a small language area, all the major research results have been reported in English, the major language of science, in international scientific journals, which have the best coverage in literature databases. Finland has a number of internationally well-known scientists but only one Nobel laureate in physiology or medicine. Ragnar Granit was born in Finland and performed a large part of his studies in the country. He later moved to Sweden and got the prize in 1967 as a Swede.

Materials and methods

We collected a list of Finnish scientists who had been active in biomedicine during the latter half of the 20th century. Data for 748 principal investigators (PIs) (mainly

professors) from universities, research institutes and industry represent well the whole nation. The PIs were identified from different sources including doctor catalogues; a listing of Finnish professors [AUTIO, 2000]; Suomen valtiokalenteri, the official yearbook of governmental positions; departmental websites; colleagues; nomination news, etc. To be included the scientist had to have published at least 10 articles in international journals and/or collected at least 100 citations. Biomedicine was considered in its widely used general definition as 'medicine based on the application of the principles and techniques of the natural sciences'. Of the studied PIs, 611 were men and 137 women. 3% were born before 1931 and 9% after the year 1969. As Finnish, we considered those born in Finland, whether active in the country or abroad, and foreigners who have spent a substantial portion of their career in Finland. There are only very few individuals who have spent most of their career abroad. They were included due to their Finnish education and their contacts to and collaborations with the Finnish scientific community. The included scientists of foreign origin have spent typically at least 15 years in Finland.

The data was collected from the Web of Science service of the Thomson Corporation. Finnish names are relatively rare. We were able to divide the productivity to individual scientists, except for 5 names which had the same initials for the second name. This was not a problem, because we were interested in the overall productivity of the country. In the case of foreign names, different means, including manual data mining and curation, and requests for lists of publications, were used to guarantee that the articles were for the investigated scientists. The data was stored in a relational database. Perl scripts, Excel macros, and MySQL queries were used for data analysis and visualisation.

Results and discussion

Analysis of Finnish biomedical research output

Records of published articles and citations made to them were collected for all the major PIs in biomedicine. The expansion of the field of biomedical research started after 1966, when there were just three medical schools; two more were established in 1972. The analysed period covers the years 1966–2000. Altogether, data for 50599 peer-reviewed articles (classified as 47303 articles, 1185 reviews, 2111 notes) was collected. The total number of citations to all the articles is 1261439 (an average of 24.9 citations per article). Of the 5900 journals covered by the Science Citation Index database, Finnish scientists have published in 3339 (57%). The database covers just a fraction of the existing journals, but only a very small number of Finnish articles have been published in other journals.

The number of articles per scientist varies from 5 to 493; the highest number of citations for a PI is 50742. The number of publications as well as the number of citations to articles published in a certain year has risen almost linearly (Figure 1a). The proportion of female PIs is 18%, their share of articles is 10423 (21%) and of citations 275761 (22%). The number of citations per article and the number of authors per article has grown steadily (Figure 1b). The analysis period ended at the year 2000. Since 1993, the citation counts decline because it takes several years after publication to accumulate citations. More details are available at http://bioinf.uta.fi/CitationAnalysis.

The average number of authors per article has risen from approximately 2.5 to a current value close to 6 (Figure 1b). At the same time, the average number of citations per article has risen only modestly, from about 24 to 31 (the results after the mid-1990's do not indicate the final citation accumulation) and thus the number of citations divided by the number of authors per paper has gone down from 10 to around 4 (Figure 1b). These are international trends indicating that today's research is performed by larger groups and consortia [CHEW, 1986; WEEKS & AL., 2004].

The distribution of the number of articles and the number of citations for each PI is shown in Figure 2, where researchers are arranged according to their total citation count. There are large variations in the numbers of articles and citations. Figure 1c shows that 115 PIs (15%) have collected 50% of all the citations, half of the citations are for articles in 71 journals, and half of the articles are published in just 167 journals. The data in the figure is organised by descending number of citations. The PIs are in descending order of their total citation numbers. 206 PIs have published half of the articles. The highest number of citations for an article is 2952 and the lowest is 0. In fact, 3775 articles have never been cited, 613 of which were published in the year 1998 or later, indicating that they had not yet been recognised. Most scientists have collected on average about 15–30 citations per article while the extremes are 2 and 148.

Recently, the h-index [HIRSCH, 2005] was introduced to measure the productivity of individual scientists. The h-index indicates the number of articles which have been cited at least h times, where h is the largest such number and is obtained by ranking the articles of a PI in descending order and finding the rank which is approximately equal to the number of citations. In our data, h-index values range from 2 to 119 (average 23.5, standard deviation 12.6). The h-index increases only very slowly along with an increasing number of publications (Figure 3). The PIs are arranged according to their total citation count. The h-index values for PIs are on average closer than, for example, average citation counts (Figures 4 and 5). The h-index for all the Finnish publications is 304.



Figure 1. Analysis of the research output of Finnish scientists. a) The yearly production of articles and the citations to articles published in certain years. b) Annual evolution of the average number of citations per article, annual evolution of the average number of authors in articles, and the annual evolution of the average number of citations per co-author.



organised by descending number of citations. Cumulative growth of citations for PIs. The PIs are in descending order of their total citation numbers. 206 PIs have published half of the articles. Cumulative growth of citations to journals. The inset indicates the numbers for the 20 most cited journals. d) Comparison of Finnish research production with the world level in 5-year periods. For each five-year period we identified 100 journals with the highest number of articles from scientists in Finland. Then, for the same journals, 500 articles (or less if not that many were published) during the middlemost year of the period were collected and the average number of citations per article was determined. The Finnish research was compared to the world level (marked as 1). The journals are arranged according to the increasing ratio.

A single parameter can never fully describe the achievements and contribution of individuals. Figure 3 indicates the h-index and a new index "v" (percentage of articles forming the h-index) for each PI. Together they better measure the contribution and recognition of scientists, and the breadth of their productivity. The h-index grows very slowly with an increase in the number of publications. The v-index indicates great variation in the proportion of highly cited articles for PIs with similar h-index values. For detailed analysis of individual PIs we use several features, including numeric values for publications, citations, average citation count, number of articles among the 100 and 1000 most cited articles by Finnish scientists, and the h- and v-indexes. In addition, we have graphs for the annual production of articles. All this data is available for each PI at http://bioinf.uta.fi/CitationAnalysis. For an example see Figure 6.



Figure 2. Analysis of article production and citation accumulation. Overall view of scientific production by Finnish scientists. The data is provided for 748 PIs (x-axis). The individual articles are on the y-axis and the cumulative citation numbers on the z-axis. The researchers were arranged according to their total citation count.



Figure 3. Analysis of the productivity of Finnish scientists. The h-index for each researcher is indicated by closed shapes and the v-index (the percentage of h-index forming articles of all articles) with open shapes. The PIs were arranged according to their total citation count. The h-index grows very slowly with an increase in number of publications. The v-index indicates great variation in the proportion of highly cited articles for PIs with similar h-index values.



Figure 4. Graph of the total citations (y-axis) for PIs having a certain h-index value (x-axis).



Figure 5. h-index values and average citation counts for each PI.



Figure 6. An example of a personal citation analysis page on the Web service. In addition to numerical information for several parameters, graphical presentations of different aspects of publication and citation records are presented. The goal of the personal data analysis is to provide a comprehensive view of the different aspects of productivity. The data is available for all the studied PIs.

Comparison to the world level

The standard of Finnish biomedical publications was compared with the world level by determining the ratio of citations of articles written by Finnish PIs to citations to all articles (Figure 1d). For each five-year period we identified the 100 journals with the highest number of articles from scientists in Finland. Then, for the same journals, 500 articles (or less if not that many were published) during the middlemost year of the period were collected and the average number of citations per article was determined. The Finnish research was compared to the world level, which was marked as 1. The journals are arranged according to increasing ratio. In each 5-year period, 30-40 of the 100 analysed journals were below world level, while correspondingly 60-70% were around or above the world average. This ratio has remained constant throughout the years. During the 1960's and 1970's a large proportion of articles were published in Finnish or Scandinavian journals but only a very small fraction since then. Articles in certain journals are substantially above the world level (up to 4 fold). The level of Finnish research in journals with the highest numbers of articles has remained at about the same level. Naturally, the volume of research has expanded several fold during the studied 35-year period.

It is noteworthy that the majority of the journals in which the level is above world standard collect on average a large number of citations. From the list of the 100 most cited articles written by Finnish scientists, 18 were published in Cell, 15 in Science, 10 in *Nature* and 6 in *Lancet* (altogether 49%). These journals are responsible for a large proportion of all the citations (Figure 1c). The articles written by Finnish scientists are consistently more cited in these journals than articles in general (Figure 7). For each 5year period we pooled the citation information for publications of Finnish scientists. For the same journals, 500 publications (or less if not that many were published) during the middlemost year of the period were collected and the numbers of citations were calculated. The total number of citations per article has grown in the journals. Articles written by Finnish authors harbour a higher number of citations than articles in general. In Figure 9 the accumulation of citations to articles in four major journals is shown. For this analysis the data was pooled into 10 year periods. The articles are organised in groups of citations (increments of 10 citations) and the cumulative growth of the citations is plotted. The articles of Finnish scientists are consistently more cited on average than articles in general. Only the very best articles of Finnish researchers have been submitted and/or accepted by these journals. On the list of the most cited articles are also journals that do not have very high impact factors. For example, the 3rd and 9th most cited articles were published in Biochemica Biophysica Acta, with 2222 and 1555 citations. The most cited articles have been published throughout the analysis period; however, there is a clear tendency for increasing numbers when entering the 1990's.



Figure 7. Citations to articles in the four journals with the largest numbers of citations. Comparison of average citation counts per article for Finnish and all authors. For each 5-year period we pooled the citation information for publications of Finnish scientists. For the same journals, 500 publications (or less if not that many were published) during the middlemost year of the period were collected and the numbers of citations were calculated. The total number of citations than articles in general. The data for all authors is shown with a continuous line and for Finnish scientists with a broken line.



Figure 8. The distribution of citations for articles published in *Nature* during the 1990's. The articles are organised in groups of citations (increments of 10 citations). For clarity, the data is shown only up to 1000 citations per article. The citation distribution is very biased and therefore journal impact factors or average citation counts cannot be used to represent individual journals.





Figure 9. The accumulation of citations to articles in the four journals with the largest numbers of citations. a) *Cell*, b) *Lancet*.





Figure 9. The accumulation of citations to articles in the four journals with the largest numbers of citations. c) *Nature*, d) *Science*.

For this analysis the data was pooled into 10 year periods. The articles are organised in groups of citations (increments of 10 citations) and the cumulative growth of the citations is plotted. The data for the 1970's is in black, 1980's in red and 1990's in blue. The articles by Finnish scientists (dashed line) are consistently more cited on average than articles in general.

Discussion

How can Finland as a small country (5.2 million people) make such an international impact? There are several reasons. Finland has a long and strong tradition in biomedical research and has been forced as a small nation to publish in international journals (in English). The country has invested heavily in research and the size of research groups has steadily grown. In the beginning of the investigated period the groups were formed by just a few individuals, whereas nowadays some groups are formed of tens of scientists.

Since the 1950's it has been the general norm that biomedical scientists have spent a post-doctoral period abroad and returned with new experiences and ideas. The Asla Fulbright program and the Fogarty International Center were crucial for the start of this practice. Finnish scientists have been well networked with colleagues in other countries as well as within Finland. This development has been further strengthened since the beginning of the 1990's by participation in EU research programs. Although only 51% of the scientists have been responsible for 75% of the citations, the whole research community has contributed to the high standard of education, and technological and scientific knowledge.

During the years 1997–2001, Finland's share of the total published articles was 0.96% and the share of citations was 1.14% and 1.1% of the top 1% highly cited publications [KING, 2004]. In another study, Finland's share of papers in science, medicine and engineering during the years 1981–1994 was 0.7% and the share of citations 0.6% [MAY, 1997]. Finland was placed 18th in the world with a share of 0.90% in a recent study of health related publications in 1992–2001 [PARAJE & AL., 2005]. Finland's R&D expenditure is 3.5% of GDP (gross domestic product). The research product of Finland, normalised by GDP or population size, was the second highest both among 15 European Union nations during 1990–1998 [HEFLER & AL., 1999] and European countries and the United States during 1994–2000 [SOTERIADES & FALAGAS, 2005]. Finnish research has previously been evaluated, based on cumulative journal impact factors, amongst the top 15–20 nations [LUUKKONEN, 1990; PERSSON & AL., 2000; SPERTI & IERARDI, 1999]. The citation ratio of all Finnish papers was 1.20 for articles published 1993–2002 among 31 countries [KING, 2004]. This is evident also in our comparison with the world level (see Figure 1d).

It would have been interesting to correlate publication output to funding. However this turned out to be impossible. Although statistical information has been collected for years, the criteria for data and classifications have changed several times. It was not even possible to estimate the funding medical schools allocated to research. It has been evaluated that in 1990 the Research Council for Health of the Academy of Finland, the Finnish Cultural Foundation, and the Sigrid Juselius Foundation funded altogether 23 m€ while the sum in 1970 was less than 4 m€ [SAXEN, 2000]. In 1994, private funding

was estimated to account for 84 m€ and government funding, including the budgets of research institutes, altogether up to 170 m€ [KARMA, 1995].

The national research output has also been investigated in Italy [SPERTI & IERARDI, 1999] and Australia [BOURKE & BUTLER, 1997; FAVALORO, 1998], but based upon journal impact factors, and in Canada by the numbers of publications [GAGNON & AL., 2000]. One might think that large numbers of articles would level out differences and inconsistencies in citations. This is naturally true to a certain extent, but the use of general journal impact factors assumes that all articles in a certain journal made a similar contribution. It is well known that, in reality, the citation numbers are very skewed (see Figures 7 through 9).

Many studies of scientific production and comparisons between countries and geographic regions [BOLDT & AL., 1999; HUSSO & AL., 2000; MELA & CIMMINO, 1998; NIKKARI, 1995; SOTERIADES & FALAGAS, 2005; SPERTI & IERARDI, 1999] have been based on a straightforward assumption of all articles being equal in respect to their contribution, despite of known pitfalls [OPTHOF, 1997; SEGLEN, 1997]. As shown above, articles have great variations in the numbers of citations. In Figure 8 are the actual citations for all the articles, notes and reviews published in one of the major, and thus very highly cited, journals, namely Nature. Of the 9829 publications during the 1990's, 175 (1.8%) have not been cited even once. 1552 (15.8%) of these papers have been cited less than 32 times, the impact factor of the journal. The data was collected at least six years after the publication date. Previous studies with a more limited dataset and time frame show the same phenomenon [OPTHOF, 1997; OPTHOF & AL., 2004; SEGLEN, 1992; SEGLEN, 1997].

Impact factor calculations take into account citations during the two years after the publication year. Our results indicate that articles by Finnish PIs, and most likely from many other major countries in biomedicine, are actually more cited than impact factors indicate. The differences in acquired citations between the major biomedicine research countries and other countries are most likely even more drastic than reported (e.g. [SOTERIADES & AL., 2006]). Our method can be applied for any geographic area or scientific field. The analysis has to be based on actual article numbers and citation counts either for named authors or, for example, for publications from a country. Even for bigger countries it is possible to identify the active researchers in a certain discipline and collect information just for them. Only by using the true article and citation counts can the real contribution of a nation be revealed.

We thank Taina Kutja, Hannu Korhonen, and Hilkka Piirilä for data collection; Jukka Lehtiniemi and Tommi O. Lehtinen for visualisation; Teemu Kivioja for some statistical analyses; Ari Huovila, Seppo Parkkila, Kalle Saksela, Olli Silvennoinen, and Tapio Visakorpi for help in identifying Finnish principal investigators; Petteri Sintonen for constructing the database; and Kathryn Rannikko for language correction. Financial support from the Medical Research Fund of Tampere University Hospital is gratefully acknowledged.

References

AUTIO, V.-M. (2000), Suomen Professorit Finlands Professorer, Professorilitto - Professorsförbundet.

- BOLDT, J., MALECK, W., KOETTER, K. P. (1999), Which countries publish in important anesthesia and critical care journals?, *Anesthesia & Analgesia*, 88 (5) : 1175–1180.
- BOURKE, P., BUTLER, L. (1997), Mapping Australia's basic research in the medical and health sciences, *The Medical Journal of Australia*, 167 (11–12) : 610–613.
- CHEW, F. S. (1986), The scientific literature in diagnostic radiology for American readers: a survey and analysis of journals, papers, and authors, *American Journal of Roentgenology*, 147 (5): 1055–1061.
- DE JONG, J. W., SCHAPER, W. (1996), The international rank order of clinical cardiology, *European Heart Journal*, 17 (1): 35–42.
- FAVALORO, E. J. (1998), Medical research in New South Wales 1993–1996 assessed by Medline publication capture, *The Medical Journal of Australia*, 169 (11–12) : 617–622.
- GAGNON, R. E., MACNAB, A. J., GAGNON, F. A. (2000), A quantitative ranking of Canada's research output of original human studies for the decade 1989 to 1998, *Canadian Medical Association Journal*, 162 (1): 37–40.
- GARCIA-RIO, F., SERRANO, S., DORGHAM, A., ALVAREZ-SALA, R., RUIZ PENA, A. & AL. (2001), A bibliometric evaluation of European Union research of the respiratory system from 1987–1998, *European Respiratory Journal*, 17 (6): 1175–1180.
- HEFLER, L., TEMPFER, C., KAINZ, C. (1999), Geography of biomedical publications in the European Union, 1990–98, Lancet, 353 (9167): 1856.
- HIRSCH, J. E. (2005), An index to quantify an individual's scientific research output, Proceedings of the National Academy of Sciences of the United States of America, 102 (46): 16569–16572.
- HUSSO, K., KARJALAINEN, S., PARKKARI, T. (2000), Suomen tieteen tila ja taso. Katsaus tutkimukseen ja sen toimintaympäristöön Suomessa 1990-luvun lopulla, Suomen Akatemian julkaisuja 6, Helsinki.
- KARMA, P. (1995), Lääketieteellisen tutkimuksen julkinen rahoitus., Duodecim, 111 (1): 73-77.
- KING, D. A. (2004), The scientific impact of nations, Nature, 430 (6997) : 311-316.
- LUUKKONEN, T. (1990), Bibliometrics and evaluation of research performance, *Annals of Medicine*, 22 (3) : 145–150. MAY, R. M. (1997), The scientific wealth of nations, *Science*, 275 (5301) : 793–796.
- MELA, G. S., CIMMINO, M. A. (1998), An overview of rheumatological research in the European Union, Annals of the Rheumatic Diseases, 57 (11): 643–647.
- MELA, G. S., CIMMINO, M. A., UGOLINI, D. (1999), Impact assessment of oncology research in the European Union, *European Journal of Cancer*, 35 (8): 1182–1186.
- NIKKARI, T. (1995), Suomen lääketieteen tutkimus bibliometrian valossa, Opetusministeriö, Korkeakouluneuvoston julkaisuja 5, Helsinki.
- OPTHOF, T. (1997), Sense and nonsense about the impact factor, Cardiovascular Research, 33 (1): 1-7.
- OPTHOF, T., CORONEL, R., PIPER, H. M. (2004), Impact factors: no totum pro parte by skewness of citation, *Cardiovascular Research*, 61 (2) : 201–203.
- PARAJE, G., SADANA, R., KARAM, G. (2005), Public health. Increasing international gaps in health-related publications, *Science*, 308 (5724) : 959–960.
- PERSSON, O., LUUKKONEN, T., HÄLIKKÄ, S. (2000), A Bibliometric Study of Finnish Science, VTT, Espoo.
- SAXEN, L. (2000), Puoli vuosisataa lääketieteen tutkimusta Suomessa., Duodecim, 116 (9): 963–970.
- SEGLEN, P. O. (1992), The skewness of science, Journal of the American Society for Information Science, 43 (9): 628-638.
- SEGLEN, P. O. (1997), Why the impact factor of journals should not be used for evaluating research, British Medical Journal, 314 (7079): 498–502.
- SOTERIADES, E. S., FALAGAS, M. E. (2005), Comparison of amount of biomedical research originating from the European Union and the United States, *British Medical Journal*, 331 (7510) : 192–194.
- SOTERIADES, E. S., ROSMARAKIS, E. S., PARASCHAKIS, K., FALAGAS, M. E. (2006), Research contribution of different world regions in the top 50 biomedical journals (1995–2002), *FASEB Journal*, 20 (1): 29–34.
- SPERTI, G., IERARDI, C. (1999), The ranking of Italy's publications in the biomedical field: an objective evaluation, *Cardiologia*, 44 (1): 51–54.
- WEEKS, W. B., WALLACE, A. E., KIMBERLY, B. C. (2004), Changes in authorship patterns in prestigious US medical journals, *Social Science & Medicine*, 59 (9): 1949–1954.