Health sciences research in Finland 2003–2005: A Comparison between the special government transfer (EVO) points and levels of evidence

Helena Tähtinen*, Päivi Rautava, and Risto-Pekka Happonen*****

* Medical Library, Turku University Library, Turku, Finland
** Dept. of Public Health, University of Turku, and Clinical Research Centre, Turku University Hospital, Turku, Finland
*** Dept. of Oral and Maxillofacial Surgery, University of Turku, and Dept of Oral Diseases, Turku University Hospital, Turku, Finland

INTRODUCTION

In Finland, the state funding (EVO) for health sciences research to health care units of five university hospital districts is currently based on the amount of special government transfer points (EVO points) produced. The number of EVO points for a single publication is based on the Impact Factor (IF) of the journal in which the article is published (Table 1). During the years 2003–2005 the value of one EVO point varied from 5.008 € to 5.914 €.

IF value	EVO points
if < 1.0	1
if = 1.0 − < 4.0	2
if ≥ 4.0	3

Table 1. IF values of the journals and corresponding EVO points

The current system favors basic research, big medical specialties and popular research fields. The journals of basic scientific disciplines, such as molecular and cell biology or immunology, tend to have much higher IF compared to that of and applied and clinical journals. This difference makes comparison between journals and research groups in different fields difficult (1). Large specialties and popular research fields have many journals, publications, and high citation activity.

The impact factor as the basis for fund allocation has been criticized because of its many flaws documented in literature. However, the use of IF has also supporters who consider it as the best available tool for research evaluation. New methods have been introduced, as listed in Taylor's article: the h-index, the g-index, the Eigenfactor, Index Copernicus, the Euro-Factor, the Reading Factor, and the PageRank etc (2). Most of them are also based on citations counts.

Impact Factor

The impact factor (IF, copyright of Thomson Reuters) was invented in 1955. It has become one of central tools in evaluating research performance in academics, universities and scientific journals (3). IF is defined as the ratio of citations in a certain year to articles published during the two preceding years divided by the total number of articles published in those two preceding years. The notion is that an article must have some value if it has been cited.

The impact factor is taken as a measure of the typical citation rate for a journal and is generally used as a surrogate metric for the quality of the research articles published in the journal of question. However, there is little or no correlation between the quality or the citation number of an individual article and the impact factor of the journal in which the article has been published (4-6).

In some disciplines the citation rate is characteristically low or it grows slowly. If the total number of citations to a journal during successive years remains about the same but the number of published articles increases, the IF value declines. This does certainly not mean that the quality of the articles has worsened (7).

The definition of a 'citable' article is also problematic. For example, excluding or including letters, reviews, editorials, observations, and news etc. effect the IF value of a journal (8). The impact factor strongly favors American journals (9). Furthermore, the proportion of the journals of smaller research fields is small in the ISI Thomson Reuters databases.

Levels of evidence

In biomedical research, the quality of publications has been evaluated by analyzing the accuracy and validity of the research methods used and the evidence of research results. Different levels of evidence have been formulated in the international literature. For example, starting from the systematic reviews of randomized controlled trials with the highest level of evidence to randomized controlled trials, controlled trials etc. (10, 11). In present study we developed a scale of three levels of evidence using the search functions and indexing methods of PubMed database (see section Methods).

AIMS

The articles (N = 10.100) published in international biomedical journals by health care units in all five Finnish university hospital districts during years 2003−2005 were analyzed. The 15 research fields producing the highest amounts of EVO points were analyzed in more detail in order to determine the levels of evidence of the articles. The aim was to find out if there is correlation between amounts of EVO points and levels of evidence of the publications.

METHODS

Publications and EVO points

Research fields of the publications in the years 2003−2005 were determined with the help of Journal Subject Terms used in PubMed Database. For, example an article published in the year 2005 in *Journal of Neurology, Neurosurgery, and Psychiatry* produced two EVO points with the IF value 3.110 in 2004. This journal is indexed in PubMed Journal Database with three subject terms - *Neurology*, *Neurosurgery* and *Psychiatry*. For this reason, each these three fields received the same two EVO points. Consequently, the amounts of publications and EVO points in this study are not exact numerical values but calculatory values.

Levels of evidence

Levels of evidence are defined somewhat differing ways in biomedical literature. They consist of four levels based on publication types, research methods and other selected criteria during research process and are described for example as levels I−IV or A−D (Oxford Centre for Evidence-based

Medicine 2009). Three upper levels were selected to be applied in this study. The levels of evidence of articles were defined with help of indexing terms and search filters in PubMed database. Determination of the publications of the fourth level would have been unreliable with the selected automatic methods.

The search terms used for levels A–C are defined below.

Abbreviations: *pt = publication type; mesh = medical subject heading; s = search filter for systematic reviews; mesh:noexp = mesh without explode function.*

Level A
Systematic reviews of randomized controlled trials defined in PubMed with the search string below:
("Randomized Controlled Trial"[pt] OR "Randomized Controlled Trials as Topic"[mesh]) AND systematic[sb]).

Level B
Randomized controlled trials defined with the search string below:
("Randomized Controlled Trial"[pt] OR "Randomized Controlled Trials as Topic"[mesh]) NOT systematic[sb].

Level C
For example of *controlled clinical trials* defined with the search string:
("Controlled Clinical Trial "[pt] OR "Controlled Clinical Trials as Topic"[mesh] OR "Cohort Studies"[mesh] OR "Case-Control Studies"[mesh:noexp] OR Comparative Study[pt]) NOT ("Randomized Controlled Trial"[pt] OR "Randomized Controlled Trials as Topic"[mesh]).

Using the term *Randomized Controlled Trials as Topic"[mesh]* in A and B search strings is not quite an orthodox way to search for this type of publications. There is some variability in PubMed regarding how it is used in indexing. This was taken in account not to miss any possible hits of randomized trials.

The first term in C search strings searches also for *longitudinal studies, follow-up studies* and *prospective studies* because of the “explode” function of the search program. The search term *systematic[sb]* is a search filter in PubMed made to search for systematic reviews in the database with numerous different terms and word combinations.

The publications of the 15 most successful research fields with highest EVO points (about 7300 publications) were included in the levels of evidence analysis. The percentage values of publications qualifying in evidence levels A, B and C of the total number of the publications in these research fields were determined. The correlation between the EVO points and percentage value of evidence levels was calculated using with Spearman correlation coefficient.

RESULTS

The 10.100 publications analysed were found to be distributed among a total of 110 different research fields. The highest amounts of EVO points were produced by large research fields with high publication activity (Fig. 1). The 15 most prominent research fields produced 49% of all publications and 52% of the EVO points during the years 2003−2005.

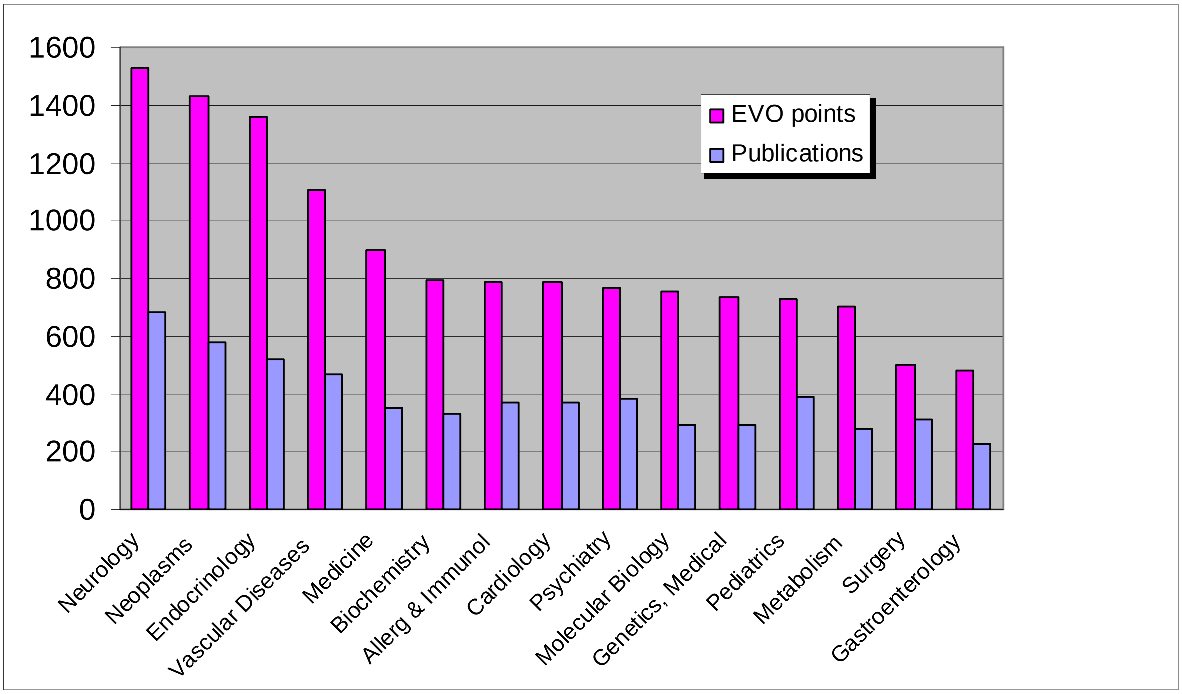

Figure 1. Amount of EVO points and number publications in the 15 most prominent research fields in 2003−2005.

A comparison between the mount of EVO points and percentage of publications qualifying into evidence levels A−C showed that there is no correlation between the two (Fig. 2; Spearman's correlation coefficient 0.03214). For example, *Gastroenterology* which ranks15th in production of EVO points is at the top in analysis of evidence levels together with *Neurology* with the highest amount of EVO points and with *Psychiatry* in 9th in production of EVO points (Fig. 2). On the contrary, the laboratory research fields (e.g. *Biochemistry*, *Molecular Biology* and *Genetics, Medical*) are among the leading fields in production of EVO points but rank low in the analysis of evidence levels.

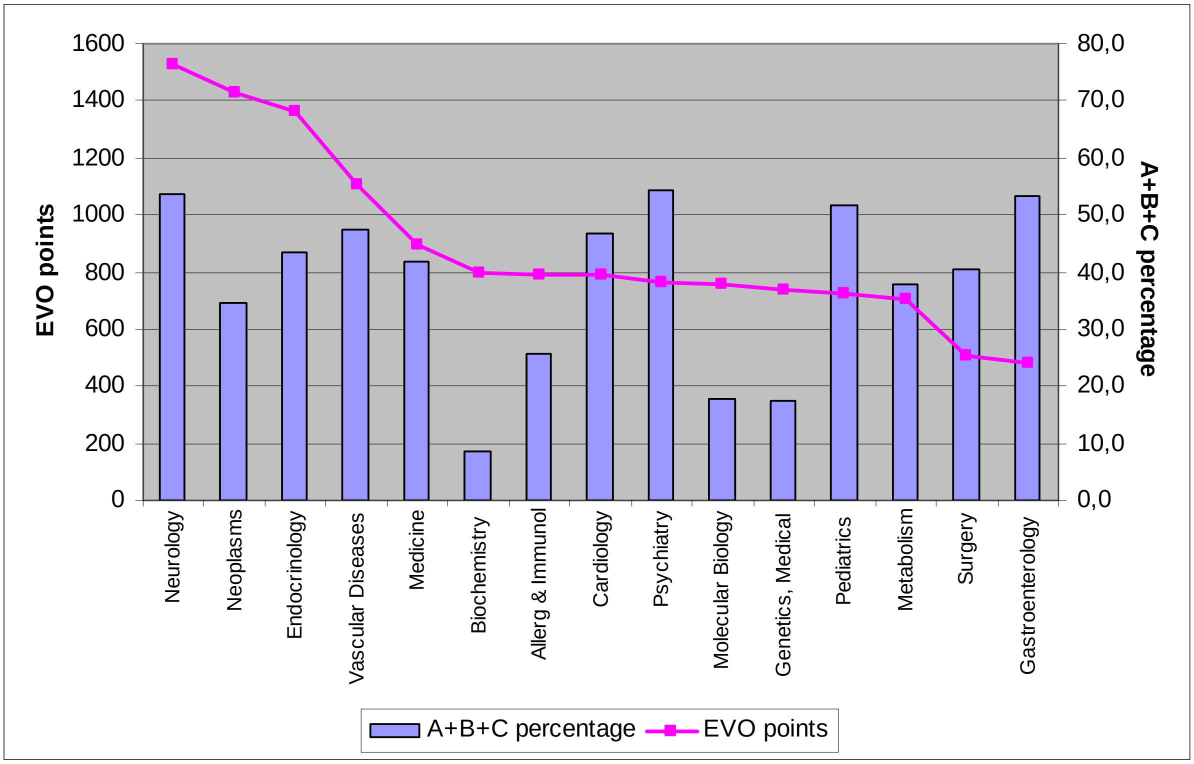

Figure 2. Comparison of the amount of EVO points and levels of evidence in the 15th most prominent research fields.

DISCUSSION

Finland is divided into five areas of responsibility of the university hospitals for organizing the expert health care services. These university hospital districts receive annually state funding to compensate the extra expenses caused by the health sciences research done. The allocation of research funding to the university hospital districts is calculated on the basis of the special government transfer points (EVO points) produced in the health care units in the area of each district. This assessment system is based on IF of the journals in which the research reports have been published. In this study, we have analyzed at national level the research output of all the five competing university hospital districts during a three year period 2003−2005 with respect to the number of publications, amount of EVO points produced, and the levels of evidence of the articles. Furthermore, the productivity of different areas of health sciences research was investigated.

The order of the four most prominent research fields was the same with regard to the amount of EVO points and number of publications: *Neurology*, *Neoplasms*, *Endocrinology* and *Vascular Diseases*. Research field *Medicine* was fifth in producing EVO points and ninth in the number of publications.

The use of PubMed Journal Subject Terms in defining the research field / fields of a single publication is not without problems. For example, the term *Medicine* represents the most respected medical journals and also other international and national series and includecs the highest number of journals in the medical field (current journals 341 in 2010). The true research field / fields of the articles published in these journals remain uncovered. Sililarly, the term *Neurology* that was the most prominent research field in producing EVO points and publications recognizes the second highest number of journal (current journals 285) in PubMed database. High number of journals recognized by a subject term has without doubt effect how the corresponding research field is doing in this kind of productivity evaluation.

Furthermore, it is not uncommon that an article is published in a journal recognized by a different subject term than is the actual research field of the study. Browsing through all publications one by one would have given more exact and reliable results. This is, however, not possible in evaluations like this dealing with a large number of publications. The Journal Subject Terms were applied in this study to define the research fields in all three analyses making their results comparable.

Analysis on the levels of evidence shows that the order of the research fields having the highest percentages of publications qualifying in the evidence levels A–C differs significantly from the order based on the amount of EVO points. Both analyses have their shortcomings. For example, significant basic research and animal studies are not at all included in the grading of evidence levels. A part of health sciences research falls out from the analysis of EVO points which is based on IF of the journals. Many journals, for example most journals of nursing sciences, do not have IF at all.

Many problems and deficiencies related to the use of IF in evaluation and comparison of research output of different scientists and institutions have been put out in the literature. Majority of articles published in a journal receive fewer citations than could be expected on the basis to the IF of the journal in question (6). Consequently, the publication forum does not prove the quality of the article. Evaluation and comparison of citations counts between different research fields do not take into consideration different citation cultures, citation growth rates and citable publication types. Year by year, the number of published articles vary which has an effect on the IF value but has nothing to do with the quality of research (7). Citations to negative or conflicting results, self-citations, citation-fishing and citation-bartending distort citation counts (12-14). Anglo-American research and popular research fields gather more citations than non-English research literature and small research fields (9). Scientists prefer often to cite publications from the most prominent countries considered more reliable than articles form countries with small research resources (15). Errors in reference lists and misspellings of names in unfamiliar languages also affect citation counts (12, 16).

There is a continuous discussion and search for the best possible and least unbiased system to evaluate research performance to be used as a fair basis for allocation of research funds.
The use of IF of the publication forum is criticized strongly as described above. Furthermore, IF favors basic research and large medical specialties and popular research areas. Many journals of health sciences have not at all IF and are thus not included in IF based analyses. The evaluation based on the levels of evidence grading treats different specialties equally but excludes basic research and animal studies.

It is concluded that neither IF or the evidence level based analysis system is as such appropriate for assessing the productivity and output of research in health care units for research allocation. In addition to the weaknesses inbuilt into both analysis systems, the human resources and economic investments used to achieve the research outcomes are not at all considered in these evaluations.

REFERENCES

1. Smith R. Beware the tyranny of impact factors J Bone Joint Surg Br. 2008 Feb;90(2):125-6.

2. Taylor M, Perakakis P, Trachana V. The siege of science. ESEP. 2008;8(1):17-40.

3. Garfield E. The history and meaning of the journal impact factor JAMA. 2006 Jan 4;295(1):90-3.

4. Not-so-deep impact. Nature. 2005 Jun 23;435(7045):1003-4.

5. Seglen PO. Why the impact factor of journals should not be used for evaluating research? BMJ British Medical Journal. 1997;314(7079):498-502.

6. Campbell P. Escape from the impact factor. ESEP. 2008;8(1):5-7.

7. Tsikliras AC. Chasing after the high impact. ESEP. 2008;8(1):45-7.

8. The impact factor game. it is time to find a better way to assess the scientific literature. PLoS Med. 2006 Jun;3(6):e291.

9. Inonu E. The influence of cultural factors on scientific production. Scientometrics. 2003;56(1):137-46.

10. CEBM > EBM tools > finding the evidence > levels of evidence [homepage on the Internet]. [cited 4/19/2010]. Available from: http://www.cebm.net/index.aspx?o=1025.

11. Lau SL, Samman N. Levels of evidence and journal impact factor in oral and maxillofacial surgery Int J Oral Maxillofac Surg. 2007 Jan;36(1):1-5.

12. Todd PA, Ladle RJ. Hidden dangers of a 'citation culture'. ESEP. 2008;8(1):13-6.

13. Lawrence PA. The mismeasurement of science. Current Biology. 2007 8/7;17(15):R583-5.

14. Agrawal AA. Corruption of journal impact factors. Trends in Ecology & Evolution. 2005 4;20(4):157.

15. Velho L. The "meaning" of citation in the context of a scientifically peripheral country. Scientometrics. 1986 01/26;9(1):71-89.

16. Kotiaho JS, Tomkins JL, Simmons LW. Unfamiliar citations breed mistakes. Nature. 1999 Jul 22;400(6742):307