Mobility, Gender and Research Quality: Combining Curriculum Vitae and Bibliometric Analysis

Ulf Sandström[1], Linkoping University, ISAK, SE-581 83 LINKOPING, SWEDEN

In recent years, Curriculum Vitaes (CVs) has been pointed to as an interesting source for social research with different approaches to the study of science. Pioneering work was done by the Research Value Mapping Program at Georgia Tech (Dietz et.al. 2000). Basically, they have shown that CVs as data for research on research have a great potential but that there are a number of methodological problems that has to be tackled. Among these are issues like availability of CVs, accuracy, coding of CVs etc. Several studies have then been presented in STS-journals (Gaughan & Bozeman 2002; Lee & Bozeman 2005; Dietz & Bozeman 2005), and other groups have connected to this research agenda (e.g. Mangematin 2000, Mangematin 2001). A small but growing research community has started to explore the possibilities of a new type of research on research with practical policy implications.

For many research issues CV-data alone might be sufficient but combining CV-data with other research databases opens up for a number of interesting research questions. This latter aspect is one of the specific features with CV data: in order to analyze many of the most interesting research questions additional data is required. This should not be interpreted as a problem but, instead, if researchers succeed in joining different databases new theoretical approaches will emerge as the ability to support and illuminate different questions with empirical data make the research more powerful. With the building of CV databases at the national and supra-national level (iniatives have been taken in Spain, Argentina, Turkey and in several other countries on the research council level) massive data will sooner or later be available and the combining of CV-data with publication databases, research grant databases and other databases is a foreseeable future.

This paper discusses the combination of CV data with publication or, we should say, bibliometric data. One of the first attempts used both kinds of data was reported by Lee & Bozeman (2005) in their important paper on research collaboration. There is a strong relationship between collaboration and productivity, but it is hard to figure out the exact relation due to a number of methodological considerations. First, there is several ways of accounting for collaboration, e.g. normal counts (all publications counted as whole counts) and fractional counts (dividing credit by the number of co-authors). Secondly, status of the researcher (e.g. professor or not professor) is of importance. Thirdly, field differences have to be taken into account.

When Lee & Bozeman (2005) used fractional counts collaborations was found to be unrelated to productivity but with normal counts collaboration was a strong predictor. Even though the "simple problem" of collaboration and productivity has been illuminated many times there are still several aspects that could be added. One the aims of this paper are to bring citations into the analysis and to look for other roads to understand the multifaceted problems of collaboration, rank, age and productivity. Another aim of this study is to test clustering analysis as a data reduction method. Is it a viable methodology for CV-studies and what results does this methodology yield? The third aim is to focus on mobility and combine that factor with citation analysis and clustering techniques. Lastly, it is interesting to include gender aspects (Cole & Zuckerman 1984; Prpic 2002) both to the productivity analysis as well as cluster analysis.

---

[1]. Corresponding author ulfsa@isak.liu.se. The author thanks Erik Sandström for generous support with programming and statistical advice.

Data and methodological issues

The primary data used in this analysis come from the curriculum vitae (CV) of 326 scientists who in 2005 were positioned as grant holders at a Swedish research foundation. The population is the principal investigators. They were granted during 2002, 2003 or 2004 as a normal grant from the foundation is for a three-year period. The average grant is approximately 60 000 Euros per year.

The CVs were collected in the time period August–September 2005 with the specific aim to perform a bibliometric study of the population. At the same time the CVs were manually coded. The coding used a template available at the Internet (www.forskningspolitik.se/cv) and developed especially for this specific dataset.

One of the advantages of this investigation is that it covers a specific area of research – oncology research, in a wide understanding. In consequence, data concern researchers active in the same field with more or less the same circumstances for the whole population. While field differences are an issue in studies using data from multiple areas and disciplines of science there are more of a common ground for comparison in this study. Though, even in this case – cancer research – it is necessary to consider differences between clinical and non-clinical research. Probably there are differences regarding output variables like number of publications and number of citations. Therefore, we will use a field-normalisation method (Glänzel 1996, van Raan 2004) in our citation analysis and the contribution of this paper is foremost to demonstrate how this can be done and that CV-analysis using this method methodologically is stronger than straight paper or citation counts.

The research area

The Swedish Cancer Society is an independent non-profit organisation. They raise and distribute money for cancer research. As the principal financier of cancer research in Sweden, the Society essentially acts as a national research council.[2]

In 2005 the Research Commission awarded around 30 million Euros to research projects, care development projects, in cancer research positions and grants. Crucial aspects when awarding grants is said to be project originality, anticipated news value and feasibility, and its link to cancer. The process of selecting research projects follows established procedures and is closely regulated. In May each year researchers submit applications describing their project's aims and its expected results, along with a cost calculation. Applications are then forwarded to eight different prioritisation committees for evaluation. The various committees' proposals are assessed in October, and in November the Research Commission makes a decision regarding which projects will be funded. Some 60 experts are used in various cancer-related areas.

Regulations stipulate that the Research Commission should follow the same principles as government research councils – such as the national Swedish Research Council – when awarding grants. The Research Commission's aim is to support the projects that represent the highest quality in the field of cancer, whatever the focus. This means that selection takes place in national competition and does not take account of the field of research or geographic location. However, there are some exceptions where targeted funds are awarded to research areas of particular urgency.

Methodological Issue 1: Quality of CVs

The problem in this investigation, as in many other CV-studies (Dietz et.al. 2000), is the shifting quality of information. Many CV:s are condensed or "truncated". In this case where the cancer

[2] This section relies on information from the CF website.

foundation asked for CVs in order to enable a bibliometric study (which indicates that you need appointments and publications) it is plausible that not all the personal information will be given by all individuals.

Table 1. List of variables with descriptive statistics

Variable	Variable type	N (=326)	Minimum	Maximum	Mean	Std. Deviation
age (in 2005)	Continous	*326*	33	76	49,48	8,697
marital status	Categorical	*326*	0	3	0,92	1,360
no of children	Continous	106	1	5	2,43	0,805
parental leave (months)	Continous	20	0	36	17,7	11,921
gender	Categorical	*326*	0	1	0,77	0,420
national orgin	Categorical	*326*	1	3	1,15	0,533
chapters and reviews	Continous	116	1	64	8,17	8,554
articles total	Continous	*326*	0	760	106,50	115,70
patents total	Continous	*326*	0	25	0,55	2,118
phd supervised tota	Continous	209	1	40	7,49	6,923
phd 1998-2005	Continous	180	1	40	4,48	4,070
postdoctors supervised	Continous	324	0	50	1,86	4,764
postdoctors 1998-2005	Continous	88	1	50	5,60	6,373
basic grants no	Continous	103	1	10	2,83	1,953
strategic grants no	Continous	35	1	5	2,06	0,873
applied grants no	Continous	46	1	7	2,07	1,162
entrepreneurial achievements	Continous	*326*	0	13	0,50	1,452
editorial appointments	Continous	*326*	0	16	0,77	2,059
trust	Continous	110	1	10	4,13	2,362
committées	Continous	104	1	20	4,84	3,184
honors&awards	Continous	94	1	28	2,48	3,155
visting professorship 1998-2005	Continous	18	1	5	1,94	1,349
year of phd	Continous	*325*	1958	2003	1987	8,6
yr of docent (associate)	Continous	262	1961	2005	1992	9,9
yr of professor	Continous	167	1968	2005	1996	7,352
mobility, no of changes	Continous	321	0	3	0,470	0,632

Note: N=326

Even a short résumé or CV can be used to get hold of important information (see Table 1) if it used together with other sources of information (Internet, CVs from other agencies etc.). The dataset consists of approximately 80 variables of which 26 is considered as interesting in this context. Not more than one third of all CVs gives information about family status and number of children. Parental leave (number of months) is the type of information of which we unfortunately are unaware of it completeness. In Swedish CVs it is not common to account for all grants received, although it seems to be of importance for many of the younger researchers. About a third of the CVs hold relevant data on research income. It should be mentioned that the CVs also give information on university, disciplinary background, research area, private grants, and all appointments, year of appointment. The latter data will be used in a section below that will demonstrate how this information can be used.

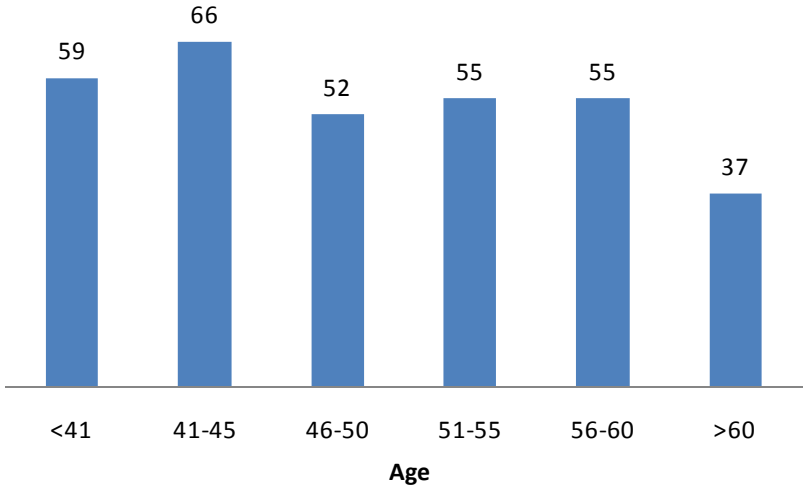

Figure 1. Distribution over age-classes, x-axis shows year of age, y-axis number of persons per class.

Figure 1 shows the distribution over age-classes. It is an interesting material as there is representation of young researchers as well as mid-aged and senior researchers. The age-groups are to a large extent of the same size and this we have to keep in mind when analysing the structure of data.

Methodological Issue 2. Coding Practice

The material was coded by one person only (the author). Therefore mistakes could be held within the boundaries of a minimum of work. The first 25 CVs was taken out and a learning process was started. After a while most of the problems were known and a strategy was outlined to handle these problems. Although many CVs report on commissions of trust, number of committees, and honours and awards it is hard to code this information. Definitions of what should be considered as commissions of trust are lacking and can be disputed, indeed. The variation for reporting this aspect among researchers is quite large. This matter can be handled in two ways: either we follow the definition in practice by each individual or we impose a more strict definition on the whole material. In this case we have followed the latter strategy.

Instead of coding each patent, entrepreneurial activity, editorial appointment, committee etc. we decided to go for a more time efficient model. As we have a parallel process of doing bibliometrics we found that it was not necessary to collect all information on each publication. Accordingly, we decided to treat all information in the same way as publications: we were interested in the period after 1997, i.e. from 1998–2005. In the coding then, we have followed this intention and notify the number of PhD:s, the number of post doctors, the number of editorial appointments, the number of visiting professorships etc. before 1998 and after 1997.

As pointed out, one of the advantages of the method used was that we did not have to code the publications. These were harvested in a parallel process using the publication list as background information for identifying all authors' articles in the ISI/Web of Science and downloading the data on each item. By giving each download the AU-name it has been possible to keep the information and align it to the CV-data in the analytical process.

Methodological Issue 3. Data reduction

Cluster analysis encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. How to organize observed data into meaningful structures are the general question and the answer could be anyone of several, e.g. factor analysis etc. We will test

cluster analysis as an exploratory tool in order to sort out the different groups in a way that the degree of association between two objects is high enough. Now, remember that cluster analysis simply discovers structures in data without explaining why they exist.

We apply a hierarchical clustering technique in SPSS with Ward's method and Squared Euclidean distance. [3] Thirteen of the variables consists of more or less complete data (age, total articles, patents, PhD's supervised, postdoc's supervised, entrepreneurial achievements, editorial appointments, year of PhD exam, and mobility) and can be used for the cluster analysis, which after several iterations create four different groups. These are found in Table 2 (next page) showing means for each cluster on several variables.

Our assumption is that researchers have an interest in giving information about their activities when they send their CV to the funding agency (or foundation). The variables shown in table 2 can be said to hold information of the type where we can expect PI:s to have an interest in contributing as complete information as possible. Some of the non-academic information (e.g. family-related issues) is missing, but the willingness to include information on items like editorials, number of PhD's supervised, number of post docs etc. seems to be very high. Regarding accuracy of CVs we have not found any notable inconsistencies in the reporting of publications.

Looking deeper into the result of the cluster analysis we find that there are four different groups. To a large extent the two first – group 1 and group 2 – are in many ways identical. The have the same age, same number of articles, and so on. The only aspect where they significantly differ is the mobility variable: the first group seem to have been looking around for new positions and new possibilities. Notable is that this type of mobility does not pay off in normal types of recognition (awards etc). The first group might be called *the mobile scientists*.

The second group might be used as reference values for contrasting the behaviour of other groups, i.e. they are what we would call "normal scientists". They are the largest cluster of researchers and fairly close to the mean of all groups. Standard deviation is high for many variables, but as "normal scientists" they seem to consist of less productive, less entrepreneurial, and have a slower career track from PhD to professorship. This is what our indicators tell us, but we might find other structures with a deeper analysis. The second group can be called *the normal scientists*.

The last two groups, group 3 and group 4, are much smaller, but they have interesting common features. In many ways the third group have the highest forms of excellence as measured by number of papers. Accordingly, they are distinguished by their low entrepreneurial activity as there is a "crowding out"-effect of patent seeking activities – entrepreneurs cannot write as many papers as researchers who concentrate on science. The third group is the oldest compared to the others and they can be characterised as highly productive, producing approximately 70 papers over the period of seven years, i.e. ten papers per year. This high productivity is explained by their supervising activities as they tend to have the highest figures for post doctors and PhDs under supervision. The group is fairly large – 37 cases – and they are extremely appreciated by colleagues and by society. Having high figures on both awards and editorial appointments they can be said to have the role of academic "watchdogs" controlling vital resources like positions (PhDs and post docs) and journals. Actually, to some extent younger researchers are the core of the third group. The third group we call *the excellence group*.

[3] Ward's method: Cluster membership is assessed by calculating the total sum of squared deviations from the mean of a cluster. The criterion for fusion is that it should produce the smallest possible increase in the error sum of squares.

Table 2. Cluster analysis (4 groups) and variable mean values per group

Cluster													
Ward Method	**Gender**	**Birthyear**	**Articl98-04**	**ArticlTot**	**Patents**	**PhD's98-04**	**Entrepre**	**Edit**	**Awards**	**Career**	**Yr of Phd**	**Mobility**	**PostDocs**
1. Mean	0,759	56,111	27,769	87,352	0,648	4,079	0,083	0,306	2,033	13,585	1988,4	1,093	1,009
N	108	108	108	108	108	63	108	108	30	53	108	108	108
Std. Dev	0,43	8,49	21,28	77,04	1,71	2,93	0,28	0,97	1,13	4,33	7,98	0,44	2,11
2. Mean	0,755	56,246	24,350	88,678	0,126	3,408	0,119	0,364	2,250	15,179	1988,8	0,000	0,455
N	143	142	143	143	143	71	143	143	32	56	143	143	143
Std. Dev	0,43	8,69	16,36	85,33	0,53	2,05	0,38	0,96	1,87	4,20	8,70	0,00	1,29
3. Mean	0,811	50,297	70,973	247,838	0,081	7,630	0,405	3,027	5,231	13,265	1981,2	0,378	9,838
N	37	37	37	37	37	27	37	37	13	34	37	37	37
Std. Dev	0,40	6,92	62,75	191,33	0,36	7,52	0,80	3,86	7,36	3,79	6,33	0,68	9,63
4. Mean	0,931	56,207	43,069	164,345	3,034	6,071	4,276	1,828	1,684	11,857	1984,7	0,621	2,207
N	29	29	29	29	29	14	29	29	19	21	29	29	29
Std. Dev	0,26	9,88	31,10	147,47	5,59	4,34	2,55	3,46	0,82	4,08	10,11	0,73	4,39
Total Mean	0,779	55,500	32,669	113,726	0,565	4,514	0,521	0,789	2,479	13,841	1987,4	0,473	1,899
N	317	316	317	317	317	175	317	317	94	164	317	317	317
Std. Dev	0,42	8,72	32,08	119,34	2,14	4,11	1,47	2,08	3,15	4,26	8,70	0,63	4,81

Lastly we have the fourth group. It is a selection of excellent researchers having a high productivity, a dedicated patenting activity, entrepreneurial activity, many editorial appointments and a fast career from PhD to full professor. The latter is shown from the career variable, which is number of years from PhD to the full professorship. They are older than average and they have had several opportunities to move around: this is obvious from mobility data which shows the number of times they have changed their affiliation during their career (not counting the post doc period). The fourth group is called *the entrepreneurial group.*

Methdological issue 4. How do we represent research quality?
As indicated above, publication data on each of the CV-authors were collected from the Thomson/ISI Internet Web of Science database. Data on all papers from the 326 researchers were downloaded using manual techniques for bibliometric identification. No automatic or matching techniques between CVs and ISI-data were used. Additionally, in order to build up reference values data on all journal articles in all subject categories where these researchers have been publishing were downloaded. Data were then gathered in a SQL database and PHP programming techniques were applied. Using the ISI UT-number as identifier we managed, to connect the author name to author names in the downloaded ISI-data. This gave analytical results from the SQL-database in the form of exact author parts of articles (author fractionalised articles).

We apply the bibliometric methods put forward by e.g. the Leiden and Leuven groups (Glänzel 1996; van Raan 2004). The methods for normalisation take all the relevant information that follows with the bibliometric data into account. Each journal in the Thomson/ISI database is assigned to one or several subject codes (e.g. microbiology, biochemistry & molecular biology etc.). Our analysis focus on Citations Per Paper (CPP) in relation to Journal Citation Score (JCS) and CPP in relation to Field Citation Score (FCS). In both cases the mean citation score for the journal set is the point of reference for the calculations (van Raan 2004; van Raan 2006). Self-citations are included in the data (at this stage) and we apply an open citation window. Articles from 1998 are compared to all articles from 1998 etc. Nederhof & Visser (2004) shows why an open window gives a methodological advantage.

The bibliometrics is based on data for 7 907 unique articles. We have performed an analysis based on citations to articles published during the period 1998–2004 and citations received until May 2007.

Results
Next, we report a complete dataset for the bibliometric analysis. There is 2,2 per cent missing data for the bibliometric analysis. Results are displayed in Table 4 [CPP/JCS =journal normalised, CPP/FCS =field normalised]

Table 4. Journal and field-normalised bibliometric values per Cluster group, and Std. Deviation (CPP/FCSm).

Ward	No. Cases	CPP/JCSm	JCS/FCS	CPP/FCSm	Std dev
(1) mobility group	105	1.17	1.33	1.49	.859
(2) normal group	141	1.11	1.28	1.40	.758
(3) excellence group	37	1.22	1.38	1.55	.823
(4) entrepreneur group	29	1.19	1.38	1.64	.643
Grand Total	**317**	**1.15**	**1.32**	**1.47**	**.782**

Note: CPP/JCS= Journal Normalized Citation Score, CPP/FCS = Field Normalized Citation Score

On the overall level, relative citations per field are 47 per cent better than world average (see column CPP/FCSm in Table 4). The citation analysis reveals that differences between groups are quite small.

In the journals where all these researchers publish they are fifteen percent more successful in gaining recognition (citations from their colleagues) than figures for world average. A closer look shows that the differences might be related to the quality of journals. While the mobility group and the normal group tend to publish in lower ranked journals, the excellence and entrepreneurial groups tend to publish in better journals or journals that yield a better result in citations. These differences are mirrored and extended in the crown indicator (CPP/FCSm), which indicate a higher citation rate or field-related performance. As can be seen from Table 4 the standard deviation is about the same for the different groups. This can be interpreted as a stable variation within the groups, which points in the direction of a systematic variation between groups also in this dimension. Certainly, we have to be careful in our interpretations. Again, it should be underlined that the variation between groups is fairly small.

There are 73 female and 252 male researchers in the material. Female researchers are about three years younger than their male colleagues. How are their respective performances when measured with field normalized methods? As in several other studies female researchers taken as a group are doing better (Long 1992; Xie & Shauman 1998, Sandström & Hällsten forthcoming). Female researchers receive a field normalized citation score of 1,53 and male researchers receive a lower figure of 1,45. Citation data are skewed and the analysis shows that the distributions are not exactly the same between the two gender groups. Trimmed mean (5 %) for females are 1,31 and for males 1,22. The range of citations performance is much wider in the male group and we can explain this with differences in publication behaviour. While male researchers tend to produce as many papers as possible of which a few will become much cited and not so few will be not cited items, female scientist are more prudent with their publications and the citation picture seem to mirror this with fewer highly cited as well as fewer non-cited articles.

Table 5. Field-normalized citation score (CPP/FCSm) per gender and cluster group

Group	Female	N	Male	N	Total	Grand total
mobility	1.58	26	1.46	79	1.49	105
normal	1.59	34	1.34	107	1.40	141
excellence	1.09	7	1.65	30	1.55	37
entrepreneur	1.16	2	1.68	27	1.64	29
Total	**1.53**	**69**	**1.45**	**243**	**1.47**	**312**

From Table 5 we can draw the conclusion that female PI:s are to a lesser extent clustered in the Excellence and Entrepreneur groups. Results indicate that female PI:s in the Mobile and Normal groups are getting higher citation scores. This might be explained by their lack of academic experience due to a lower age, but, of course, probably it reflects hierarchical structures within the academic community.

Collaboration and productivity

The research problem involved when we investigate the relations between collaborations and productivity concerns the fact that research policy analyst seldom have had access to data on actual funding. Probably the distribution of research funding is uneven and a few in the upper end of the spectrum holds a good share of the money while the majority have to live with ordinary funding arrangements. There is a *matthew effect* in research policy: if you have funding from one source your chances of getting funded from other sources will probably be higher. The size of the research group differs to a large extent according to the level of funding and if you have large resources you even might invest in facilities which open up possibilities for more collaboration with other researchers wanting to use your equipment. A plausible hypothesis would be that some of the “best” researchers are capable of assembling large resources and that this is a forgotten explanation as to why some are

more “productive” than others. This is dealing with spurious relations as we cannot know whether the productivity was there before the funding, but it is more or less an empirical question.

CV-data does not give actual research funding and composition of resources. Just as Lee & Bozeman (2005) our data on this aspect is too weak to be put into the analysis. To some extent we could use our cluster analyses as a proxy for this aspect. Looking closer to the analysis there are several indications that the excellence group, which seem to be more fortunate in several respects, could be seen as a more resourceful group. They have more PhD-students, they have more post docs, they receive a higher number of awards, and consequently they have a high number of publications. This is confirmed by an analysis of cluster groups in relation to publications, both normal and fractional counts. Table 6 shows that there are big differences between groups in their productivity and that the Excellence group have the highest number of fractional counts and normal counts as well. Their productivity is almost doubled in relation to the productivity of the first and second group.

Table 6. Normal Normal counts and fractional counts per cluster group (mean values per author during 1998–2004).

Group	**Normal count** PROF	**Frac count** PROF	**N** PROF	**Normal count** NON-PROF	**Frac count** NON-PROF	**N** NON-PROF
(1) mobile	37,40	8,6	52,0	18,49	3,9	53,0
(2) normal	30,82	6,4	57,0	18,70	4,1	85,0
(3) excellence	69,68	15,3	34,0	71,67	17,4	3,0
(4) entrepreneur	51,57	11,1	21,0	19,75	4,4	8,0
Totalt	**43,62**	**9,5**	**164,0**	**19,74**	**4,3**	**149,0**

Note: Articles are counted during 1998–2004.
Pearson correlation between Normal and Frac count is 0.93.

Table 7 goes further into the differences according to the number of years since PhD was received. While professors have a higher number of papers per period, i.e. a higher productivity, the non-professors in the material, persons with a lower academic status, seem to have almost the same productivity no matter of how long time since they received their PhD. Another aspect is that the productivity does not always implicate a higher number of field normalized citations per paper. The so called crown indicator is more or less unrelated to the productivity. The best papers seem to be written by younger researchers no matter whether they are professors or not. This seem to stable even when controlled for by gender. The burden of high productivity seems to be that your relative citations become a bit lower. In conclusion, empirical data supports the view that productivity is related to professorial status, but at the same time there is evidence pointing in the direction that innovative research tend to come with younger researchers who received their PhD:s during the 1990s.

Table 7. Fractional counts per group of years since PhD (mean values per author during 1998–2004).

Yrs since PhD	Prof frac papers	non-Prof frac papers	Prof CPP/FCSm	non-Prof CPP/FCSm	N Prof	N non-Prof
B (>29 yrs)	12,5	3,8	1,17	0,82	29	1
C (25-29)	11,7		1,39		31	
D (20-24)	8,0	4,2	1,43	1,03	42	6
F (15-19)	8,9	5,5	1,37	1,20	36	20
G (10-14)	7,1	4,7	1,68	1,28	27	53
H (5-9)	7,5	3,8	>3,00	1,68	2	65
I (>5)		2,2		1,82		8
Total	**9,5**	**4,3**	**1,46**	**1,46**	**167**	**153**

Gender makes a difference in productivity (data not shown here). Female researchers have significantly lower number of papers in their respective groups, but it should be noted that among professors there are slightly smaller variations between male and female researchers.

Researcher mobility

In this section we will analyse the pattern of mobility within this specific area of research (oncology). CVs are probably the best way to study mobility issues. They give information on all shifts from one place to another and from one position to the other. Data on these aspects seem to be more complete than other aspects although we cannot be sure that every move during a career is accounted for by everyone. In our data we have 1.472 moves. Not all of these represent mobility from one institution to the other as we also record all changes in position (appointments). The mobility between institutions can be counted to a sum of 594, which indicates that the average researchers have almost two inter-institutional moves during their career. Of course, as discussed above the distribution of mobility is uneven and some are “mobile” researchers and others are “immobile”.

Table 8 shows the mobility patterns from one institution (row) to another institution (column). Although the post doc experience is of interest in this context we have decided not to include that information in Table 5. Mobility due to the post doc can be considered as a form of “institutionalised” mobility, valuable as such but with a specific logic. Each instance of mobility is registered as a change from the row to the column. There are 15 moves from UU (Uppsala University) to KI (Karolinska Institute). There are 50 moves from UU and only 31 researchers coming o Uppsala. We could interpret this figure in different ways, but it seems obvious that Uppsala is a supplier of competence, while KI is a demanding “machine”. There are 69 moves to KI and 38 from, i.e. a positive surplus. The matrix shows that large universities are generally demanders and to a lower extent suppliers of people to the system. Lund University (LU) and Gothenburg University (GU) both have a positive surplus while Umeå University (UMU), Stockholm University (SU) and KTH (Royal Institute of Technology) have a negative. The Institute Sector seems to be a neutral player in this field of research. They receive as many as they supply which indicates that they have a stable situation regarding financing of research.

Table 8. Matrix showing mobility between institutions (all data not shown)

FROM/TO	→KI	→UU	→LU	→UMU	→INSTITUTE	→GU	→USA	→OTHER	→SU	→LIU	→UK	→FINLAND	→PRIVATE	→SLU	→GERMANY	→FRANCE	→KTH	→DENMARK	SUM	% MOVES FROM
NO. OF PEOPLE	**140**	**75**	**68**	**45**	**39**	**36**	**28**	**26**	**24**	**15**	**8**	**8**	**7**	**5**	**5**	**3**	**3**	**2**		
KI →	(490)	6	2	4	7	2	1	6	4	2	0	0	1	0	0	0	1	0	**38**	**36%**
UU →	15	(224)	7	1	9	0	8	3	1	1	0	0	2	2	0	0	0	0	**50**	**62%**
LU →	2	1	(194)	0	1	3	2	2	0	2	2	0	2	0	0	0	0	2	**19**	**36%**
UMU →	5	2	1	(120)	3	1	3	2	2	0	0	0	1	0	0	1	0	1	**25**	**54%**
INSTITUTE →	10	6	5	4	(36)	2	3	1	1	0	1	0	0	0	0	0	0	0	**34**	**50%**
GU →	2	0	1	1	0	(124)	1	0	0	0	0	0	0	0	0	0	0	0	**6**	**33%**
USA →	4	6	3	1	4	1	(8)	2	2	0	1	0	0	0	0	0	1	0	**26**	**54%**
OTHER →	9	3	4	1	2	1	1	(12)	0	1	0	0	0	0	0	0	0	0	**22**	**54%**
SU →	6	1	1	3	0	0	2	1	(50)	1	1	0	0	0	0	0	0	0	**16**	**57%**
LIU →	3	0	0	1	0	1	0	1	0	(28)	0	0	0	0	0	0	0	0	**6**	**46%**
UK →	3	0	2	1	1	0	0	0	1	0	(0)	0	0	0	0	0	0	0	**8**	**62%**
FINLAND →	1	0	1	0	4	0	0	1	0	0	0	(20)	0	0	1	0	0	0	**8**	**100%**
PRIVATE	1	2	2	1	1	0	0	0	0	0	0	0	(6)	0	0	0	0	0	**7**	**54%**
SLU	1	1	0	0	0	0	0	0	0	0	0	0	0	(16)	0	0	0	0	**2**	**50%**
GERMANY →	1	1	2	1	0	0	0	0	0	0	0	0	0	0	(2)	0	0	0	**5**	**83%**
FRANCE →	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	(0)	0	0	**2**	**67%**
KTH →	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	(2)	0	**3**	**60%**
DENMARK →	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	(0)	**3**	**50%**
SUM	**69**	**31**	**34**	**21**	**34**	**12**	**22**	**19**	**12**	**7**	**5**	**0**	**6**	**2**	**1**	**1**	**2**	**3**		

Source: Data from 326 CVs. Figures show number of moves between institutions.

It is possible to visualize the information in Table 8. We use CiteSpace and normalize the strength of relations according to the size of each institution and the number of moves between institutions. Figure 2 illustrates the information showing the size of the institution and the relative position in the landscape. What is interesting in Figure 2 is that Lund University (LU) has a pattern of mobility that is more related to other countries than to other Swedish universities.

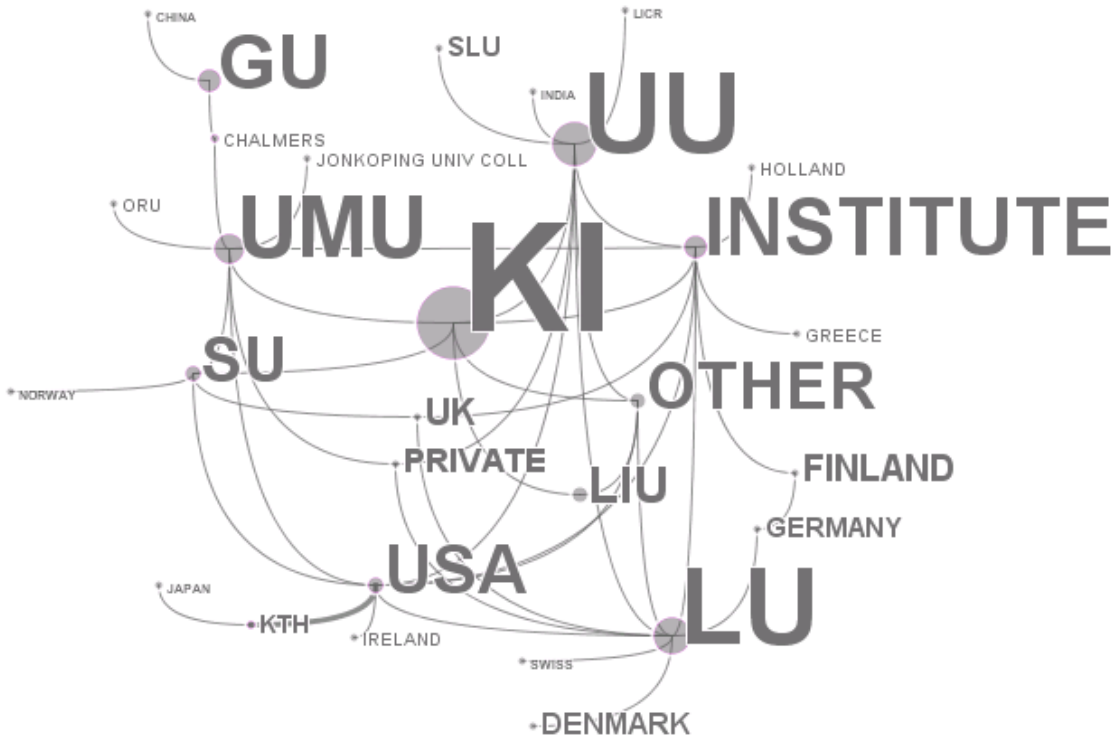

Figure 2. Pattern of mobility in Swedish oncology research

Searching for trends in mobility

In this section we intend to match the different approaches we have used. Our question is whether it is possible to utilise the data for a study of traits of mobile persons and the direction in their mobility. We are dealing with the Swedish research landscape and the competition between universities. In our understanding some of the institutions are winners and some are losers. Is it possible to put figures on this? In order to achieve this we propose a switch from the analysis in the last section where we used moves between institutions to moving persons instead. In Table 9 is displayed the result of this analysis. In the table is shown the number of persons moving and the figure of performance (CPP/FCSm) for these researchers. The table should be read in this manner: A group of researchers start their career at KI and they are still there or have been on a career move to other places. This group of 73 persons has a CPP/FCSm of 1.49, i.e. 49 per cent above average. Uppsala (UU) loses a group of 15 researchers that now have an appointment at KI. This group is considered as high performers with a CPP/FCSm of 1.59. Uppsala has a large group that stays or come back to UU and these “local heroes” are also very good performers (1.57).

In the bottom row line is found the average CPP/FCSm for all persons that have moved to KI etc. In the last column we find the other end of the spectrum: all persons moving from KI etc. Note that these figures report all the mobility to and from these universities. Comparing the means of citation rate makes it possible to draw some preliminary conclusions. Karolinska Institute is the big winner in the competition over human capital in Swedish oncology research. Uppsala loses people with 1.50 and receive a human capital with 1.30. The same apply for Umeå. Lund University has a positive exchange and this is also the case for Linkoping and Stockholm University. Figures for “local heroes” are often markedly lower than for the mobile group, except for Uppsala.

Table 9. Mobility and Citation Rate

	→ KI	→ UU	→ LU	→ UMU	→ GU	→ SU	→ LIU	AVG CPP/FCSm →
NO. OF PEOPLE	**140**	**75**	**68**	**45**	**36**	**24**	**15**	
KI →	[1.53 (73)]	1.31 (3)	1.32 (2)	1.64 (1)	2.08 (1)	1.15 (2)	0.93 (1)	1.41 (15)
UU →	1.59 (15)	[1.57 (28)]	1.15 (3)	1.08 (1)	1.58 (2)	-	1.73 (1)	1.50 (29)
LU →	-	-	[1.22 (43)]	-	0.72 (2)	-	1.4 (2)	1.06 (4)
UMU →	-	1.98 (2)	1.02 (2)	[1.35 (20)]	1.39 (2)	3.55 (1)	-	1.77 (12)
GU →	1.98 (1)	-	-	0.64 (1)	[1.16 (22)]	-	-	1.68 (3)
SU →	0.90 (4)	0.75 (1)	1.64 (1)	1.66 (2)	-	[1.03 (5)]	-	1.23 (10)
LIU →	-	-	0.90 (1)	-	-	-	[1.29 (5)]	0.90 (3)
→ AVG CPP/FCSm	1.68 (45)	1.45 (10)	1.40 (19)	1.35 (8)	1.60 (9)	1.48 (6)	1.36 (4)	

Note: Row summary and Column summary report data not shown in the table.

Cluster Groups and Mobility

In the final section we will try to integrate cluster analysis and mobility analysis. Is there a typical path for the different cluster groups? There are, of course, many paths of mobility between Swedish universities or between a foreign country and a Swedish university. In this analysis we will focus on the most typical traits for each cluster group. As there are so many paths, each one unique for that person, but seen as a group there might be some features of mobility figures that could be of interest.

Analytically it is hard to get hold of all dimensions that should be included in this exercise. There are a lot of mobility and many different variables that might be of interest. In many ways a qualitative analysis could be the best way to account for all the variation in the material. *Group 1* has been characterized as the **mobility** group. When we dig deeper into the data we find that there is a large group of foreign scientist in this group with a PhD from abroad. They tend to move around in Europe before they find themselves a permanent position.

Going further to *Group 2* we find that there are a few typical paths. This is the **immobile** group. Their most typical career path starts in Lund or Umeå and it ends in the same places. Of course, this is a disappointing result, but probably a reflection of the low mobility of researchers in Sweden. Now, remember that we are investigating the mobility in relation to where researchers get their PhD and where they end up as professors.

What is typical for *Group 3*? They are the most **prestigious** of researchers, characterised by their excellence, seniority, and their "watchdog" experiences. Starting their career at Karolinska or Uppsala, they tend to end up at their mother university, which might be explained by the resources offered by that university. Actually, a short time period between PhD and professorship does not seem to be connected to an overall pattern of mobility: they start at a university and ends up at the same university. Of course, this group might have a short period at another university but after this experience they are recruited back to their home university. And it all happens during a short time period.

Lastly, we have the mobility behaviour of the **entrepreneurial** group. First, we should notice that the post doc experience differentiate between groups. In general, about half of each group have done a post doc abroad, mainly in the United States. The entrepreneurial group are deviating from this quite

strong rule: they have to a larger extent, about 75per cent, been active abroad during their post doc. They tend to stay longer abroad, but when they come back to Sweden they find themselves a position at the best institutions, i.e Uppsala University and Karolinska Institute.

The conclusion of this analysis is that in general it is not possible to sort out the typical mobility pattern within cluster groups. As soon as we want to use large numbers we always end up with the large group of "local heroes" –researchers that receive their PhD at a university and end up as professors at the same university.

Conclusion

The main contribution of this paper is that citation analysis have been included and tested on a large dataset of CV:s. Using field normalized citations and papers during a specific period of time has given new insights and new perspectives to the problems of collaboration and productivity. A first result is that gender is related to productivity in a double way: female scientist may be less productive in both normal and fractional paper counts, but their impact in field normalized citations are better than their male colleagues. This is the paradox of gendered productivity patterns which seem to be relevant in many areas of research. My conclusion is that this should be the focus of more analysis and comparative research. If we do not understand the effects of horizontal division of labour we will probably fail to understand how different areas of research are developing.

By introducing cluster analysis of the CV data we have been able to pinpoint some of the patterns underlying this area of cancer research in Sweden. The four groups of researchers – Mobile, Normal, Excellence and Entrepreneur – reveal some differences when it comes to certain important variables like gender and impact of research. Female scientists are overrepresented in the first two groups and under represented in the latter groups. One implication of this is that relatively fewer female researchers have the type of resources that are assembled by their male colleagues in the Excellence and Entrepreneur groups. Well-known *Matthew Effects* probably lies behind and can explain why this is the case. As science study researchers we might point out that females with their tendency to lower productivity but higher citation rates will have to find ways of translating their revealed excellence into ways of assembling resources. This they will have to do in relation to research foundations and research council that up to now predominantly have been interested in number of publications (Brouns 2000; Sandström & Hällsten 2008). With a pessimistic look at things female researchers will not have the ability of assembling resources for their research. Instead, they will have to align with older and maybe not so innovative male colleagues that control the funding resources and still are the main receivers of main resources of funding.

As stated above female researchers are over-represented in the mobile and normal groups of our clusters. Within these groups they seem to perform better research with more citational reputation. Collaborative research is productive but less collaborative research is relatively more cited work. This result goes against many of the propositions in earlier research. Lee & Bozeman (2005) hold the door open for conclusions like these as they do not have a qualitative indicator, and my conclusion is that these fundamental questions have to be further investigated. One way of taking the research agenda to new aspects is the contribution of the mobility study in this paper. We have not been able to connect the gender variable to this mobility analysis, but the results point in an interesting direction. Mobility between Swedish universities seems to be a process of searching for good researchers. Therefore, mobility can probably be understood as to a large extent determined by the quality of researchers. These are aspects that will be explored in future papers.

References

Brouns, M (2000) The Gendered Nature of Assessment Procedures on Scientific Research Funding: the Dutch Case. Higher Education in Europe 25:193-199.

CiteSpace, available at Internet (http://cluster.cis.drexel.edu/~cchen/citespace/).

Cole JR, Zuckerman H (1984). The Productivity Puzzle: Persistence and Change in Patterns of Publication of Men and Women Scientists. In Maehr & Steinkamp, eds., Advances in Motivation and Achievement pp. 217–256. Greenwich: JAI Press.

Dietz JS & Bozeman, B (2005) Academic Careers, Patents, and Productivity: industry experience as scientific and technical human capital. Research Policy 34: 349–367.

Dietz JS, Chompalov I, Bozeman B, Lane EO, Park JW (2000) Using the curriculum vita to study the career paths of scientists and engineers: an exploratory assessment. Scientometrics 49: 419–442.

Gaughan M & Bozeman B (2002) Using Curriculum Vitae to compare some impacts of NSF research grants with center funding. Research Evaluation 11: 17–26.

Glänzel W (1996). The need for standards in bibliometric research and technology. Scientometrics 35:167–176.

Long JS (1992) Measures of sex differences in scientific productivity. Social Forces 71 (1): 159-178

Mangematin V (2000) PhD job market: professional trajectories and incentives during the PhD. Research Policy 29 (6): 741-756.

Mangematin V (2001) Individual careers and collective research: is there a paradox? International Journal of Technology Management 22 (7-8): 670-675.

Nederhof AJ, Visser MS (2004) Quantitative deconstruction of citation impact indicators - Waxing field impact but waning journal impact. Journal of Documentation 60 (6): 658-672 2004

Prpic K (2002) Gender and productivity differentials in science. Scientometrics 55 (1): 27-58

Sandström, U & Hällsten, M (2008) "Research Quality and Diversity of Funding: A Model for Relating Research Money to Output of Research", Scientometrics (forthcoming)

Sandström, U & Hällsten, M (2008) Persistent Nepotism in Peer Review. Scientometrics vol 74, no2.

Van Raan AFJ (2004). Measuring Science. [Eds.] Moed, Glänzel & Schmoch. Handbook of Quantitative Science and Technology Research. Dordrecht: Kluwer Acad Publ.

van Raan AFJ (2006) Statistical properties of Bibliometric indicators: Research group indicator distributions and correlations. Journal of the American Society for Information Science and Technology 57 (3): 408-430

Xie Y & Shauman KA (1998) Sex differences in research productivity: New evidence about an old puzzle. American Sociological Review 63 (6): 847-870