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

Ulf Sandström

[Accepted for publication in *Research Evaluation* 2009]

This paper discusses the potentials of combining CV-studies with advanced bibliometrics. Based on data from 326 curriculum vitae within one broad medical subject area we perform a cluster analysis of CV data. Data reduction produces four different groups of scientists: 1) mobile, 2) normal (immobile), 3) excellent and 4) entrepreneurial. Results indicate that female researchers perform slightly higher impact papers (field normalized citations) and that mobility is related to performance, but in a mediated and complex way. Gender seems to be unrelated to the mobility of researchers and this is mainly due to the performance of immobile female researchers. While it is clear that the most mobile and the least mobile researchers represent opposites also in performance we have to acknowledge that for the large majority, with a low and medium mobility, there is no linear pattern of performance. The paper points at a double process where there are on the one hand selection processes at universities picking out "the winners" and on the other hand self selection processes where researchers enhance their own performance by being mobile.

In recent years, curriculum vitae (CVs) have been identified as an interesting source for research with different approaches to the study of science and technology. Pioneering work was done by the Research Value Mapping Program (Dietz et.al., 2000) showing that CVs have great potential as data for research on research but that there are a number of methodological problems that have to be tackled. These include issues like availability of CVs, accuracy and coding of CVs. Several CV-studies have been presented (Dietz & Bozeman, 2005; Gaughan & Robin, 2004; Gaughan & Bozeman, 2002; Lee & Bozeman, 2005), and other groups have joined this research agenda (e.g. Mangematin 2000; Mangematin 2001; Cañibano 2008).

This paper discusses the combination of CV data with bibliometric data. A first attempt to use both 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 relationship due to a number of methodological considerations. First, there are 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, researcher status (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, collaboration 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 discussed intensively, there are still several aspects that could be added. One aim of this paper is 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 is to test clustering analysis as a data reduction method. A 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) in the analysis.

Data and methodological issues

The primary data used in this analysis come from the curriculum vitae of a population consisting of all senior scientists who were positioned as grant holders at a Swedish research foundation by June 2005, in all 326 researchers (N=326). Concerning the research area it can be described in terms of medical, both preclinical and clinical, research. The population consists of principal investigators (PI:s). These PI:s received grants during 2002–2004; a normal grant lasts for a three-year period. The average grant was approximately 60 000 Euros per year.

The CVs were collected in the period September–October 2005 with the specific aim of performing a bibliometric study of the population. The manual coding used a template available on the Internet (www.forskningspolitik.se/cv), which was developed especially for this dataset.

While field differences are an issue in studies using data from multiple areas and disciplines of science, there is more common ground for comparison in this study. However, even in this case – cancer research – it is necessary to consider differences between clinical and non-clinical research (Narin & Hamilton, 1996). Probably there are differences regarding output variables like number of publications and number of citations. Therefore, we use a field-normalization method (Glänzel, 1996, van Raan, 2004) in our citation analysis and the main contribution of this paper is to demonstrate that CV analysis using this method is stronger methodologically than straight paper or citation counts.

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 CVs are condensed or “truncated”. The collection of CVs was done in order to enable a bibliometric study, which indicated that appointments and publications were needed. Accordingly, not all the personal information was given by all individuals.

But, even a short résumé or CV can be used to get hold of important data (see Table 1) if it is used together with other sources of information (Internet, CVs from other agencies etc.). The dataset consists of approximately 80 variables of which 26 are considered as interesting in this context. No more than one third of all CVs give information about family status and number of children. Parental leave (number of months) is the type of information where we unfortunately are unaware of the completeness of data. Swedish CVs does not commonly report all grants received, although it seems to become important for the younger population. About a third of the CVs hold relevant data on research income. Of course, there are information on university, disciplinary background, research area, and all appointments, with the year of appointment. The latter data will be used in a section below that will demonstrate how this information can be used.

Table 1. List of variables with descriptive statistics

Variable	Variabel type	N	Min	Max	Mean	Std.Dev.
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 (month)	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 tot	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	Continous	103	1	10	2.83	1.953
strategic grants	Continous	35	1	5	2.06	0.873
applied grants number	Continous	46	1	7	2.07	1.162
entrepren achievement	Continous	*326*	0	13	0.50	1.452
editorial appointments	Continous	*326*	0	16	0.77	2.059

commissions of trust	Continous	110	1	10	4.13	2.362
committees	Continous	104	1	20	4.84	3.184
honors & awards	Continous	94	1	28	2.48	3.155
visiting professorships	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	Continous	321	0	3	0.470	0.632

Note: N=326, most variables are counted as number of.

Methodological Issue 2: Coding Practice

The material was coded by the author. The first 25 CVs were examined and a learning process was started. After a while most of the problems were known and a strategy was outlined to handle these problems.

Instead of coding each patent, entrepreneurial activity, editorial appointment, committee etc. it was decided to use 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.

Methodological Issue 3: Data reduction

Cluster analysis is a generic term for a data reduction based on multivariate data analysis (Tjissen & de Leuuw, 1988). The grouping of objects of similar kind into categories is done with a hierarchical clustering technique in SPSS (Ward's method). Thirteen of the variables consist of more or less complete data (age, total articles, patents, PhDs supervised, post docs supervised, entrepreneurial achievements, editorial appointments, year of PhD, and mobility) and the following analysis is based on these variables. Four different clusters were identified (see Table 2).

Looking deeper into the result of the cluster analysis we find that there are four different groups. To a large extent the first two – group 1 and group 2 – are identical. They are the same age, have the same number of articles, and so on. But, in one aspect they differ significantly: mobility. In the process of building a career the first group seems to have been looking around for new positions and new possibilities. It is notable 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". This is the largest cluster of researchers and fairly close to the mean of all groups. The standard deviation is high for many variables, but as "normal scientists" they seem to consist of less productive, less entrepreneurial workers, and have a slower career track from PhD to professorship. The second group can be called *the immobile scientists*.

The last two, group 3 and group 4, are smaller in size, but they have interesting common features. In many ways the third group has 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 docs and PhDs under supervision. The group is fairly large – 37 cases – and they are very highly 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. The third group is called *the excellence group*.

The fourth group is a selection of excellent researchers with high productivity, a dedicated patenting and entrepreneurial activity, many editorial appointments and a fast career from PhD to full professor. The latter is seen from the career variable, which is the 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 show 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*.

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

Cluster													
Ward Method	Gender	Yr of Birth	Articles98-04	ArticlTot	Patents	PhD's 98-04	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

Methodological issue 4: Research impact

As indicated above, publication data on each of the 326 researchers were collected from the Thomson/ISI Internet Web of Science database. Data on all papers were downloaded using manual techniques for bibliometric identification. 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. Using the ISI UT number as the identifier we connected the author name to author names in the downloaded ISI data. This gave analytical results from a 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 (Schubert et al., 1988; 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 sub-fields (e.g. microbiology, biochemistry & molecular biology etc.). Our analysis focuses 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 and we apply an open citation window. Articles from 1998 are compared to all articles from 1998 etc. Nederhof & Visser (2004) have shown why an open window gives a methodological advantage.

In all, the 326 PI:s produced 7 907 unique articles. The bibliometric analysis is based on citations up until May 2007 to articles published during the period 1998–2004.

Results

The CV material is interesting because the distribution over age-classes is fairly even. There are 50-60 in each of six age-classes (<41; 41-45; etc), so the representation of young researchers as well as mid-aged and seniors is satisfactory.

Table 3. Journal and field-normalised bibliometric values per cluster group, and Std. Deviation (CPP/FCS).

Ward	No. Cases	CPP/JCS	JCS/FCS	CPP/FCS	Std dev
(1) mobility group	105	1.17	1.33	1.49	.859
(2) immobile 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

At the overall level, relative citations per field are 47 per cent better than world average (see Table 3). The citation analysis reveals that differences between the groups are quite small; a variation around ten per cent is probably not significant. In the journals where all these researchers publish, they are fifteen percent more successful in gaining recognition (citations from their colleagues) than the global averages. A closer look shows that the differences might relate to the quality of journals. While the mobility group and the immobile group tend to publish in lower ranked journals, the excellence and entrepreneurial groups publishes in higher ranked journals or journals that yield a better result in citations. These differences are mirrored and extended in the field indicator CPP/FCS, which reveal a higher citation rate or field-related performance. As shown in Table 3 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.

There are 73 female and 252 male researchers in the data set. Female researchers are, on average, about three years younger than their male colleagues. What 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 slightly 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. The trimmed mean (5 %) for females is 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, a few of which will become much cited and not so few that will be zero cited, female scientists are more prudent with their publications and the citation picture seem to mirror this with fewer highly cited as well as fewer not cited articles.

From Table 4 we draw the conclusion that female principal investigators are clustered to a lesser extent in the Excellence and Entrepreneur groups. The results indicate that female PIs in the Mobile and Immobile groups are getting higher citation scores. This might be explained by their lack of academic experience due to a lower age, but, of course, it probably also reflects hierarchical structures within the academic community.

Table 4. Field-normalized citation score (CPP/FCS) per gender and cluster group

Group	Female	N	Male	N	Total	Grand total
Mobile	1.58	26	1.46	79	1.49	105
Immobile	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**

Collaboration and productivity

A thorough investigation of the relations between collaboration and productivity would need full data on actual funding (grants). Probably, the distribution of research funding is uneven and a few in the upper end of the spectrum hold a good share of the money while the majority have to live with ordinary funding arrangements (Cesaroni & Gambardella, 2003). There seem to be a Matthew Effect (Merton, 1968) in research policy: if you have funding from one source your chances of getting funded from other sources will be higher. Research group size differs considerably according to the level of funding and if you have large resources you might even invest in facilities that 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.

Our dataset does not include actual research funding or the composition of resources. Data on this aspect are often too weak to be used for a detailed analysis (cf. Lee & Bozeman, 2005). Instead, we propose that our cluster analyses could act as a proxy for this aspect. Looking more closely at 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 5 shows that there are big differences first, between status groups and second, between cluster groups in their productivity. Among these different groups the Excellence group have the highest number of fractional counts and normal counts as well. They are almost twice as productive as the first and second groups.

Table 5. Normal counts and fractional counts per cluster and status group (mean values per author during 1998–2004).

	Professors			Others		
Group	Normal count	Frac count	N	Normal count	Frac count	N
(1) Mobile	37.40	8.6	52	18.49	3.9	53
(2) Immobile	30.82	6.4	57	18.70	4.1	85
(3) Excellence	69.68	15.3	34	71.67	17.4	3
(4) Entrepreneurs	51.57	11.1	21	19.75	4.4	8
Totalt	**43.62**	**9.5**	**164**	**19.74**	**4.3**	**149**

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

Number of years since receiving PhD makes a difference (Table 6). While professors have a higher number of papers per period, i.e. higher productivity, the non-professors in the material, persons with a lower academic status, seem to have almost the same productivity no matter the length of time since they received their PhD. This illustrates the hypothesis that you need a professorial status in order to assemble resources for productivity. Another aspect is that productivity does not always imply a higher field normalized citation score. The so called “crown indicator” is more or less unrelated to productivity. More highly cited papers seem to be written by younger researchers no matter whether they are professors or not. This is stable even when controlled for by gender. The burden of high productivity seems to be that your relative citations become a bit lower, which is reasonable as professors have to collaborate with all their staff. In conclusion, empirical data supports the view that productivity is related to professorial status, but at the same time there is evidence indicating that the most innovative research originates from younger researchers who received their PhDs in the 1990s.

Gender makes a difference in productivity (data not shown here). Female researchers have a 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.

Table 6. Fractional counts per group of years since PhD (mean values per author during 1998–2004) and field normalised citations score.

	Prof		Others		N	
Yrs since PhD	**Frac P**	**CPP/ FCS**	**Frac P**	**CPP/ FCS**	**Prof**	**Others**
B (>29 yrs)	12.5	1.17	3.8	0.82	29	1
C (25-29)	11.7	1.39			31	
D (20-24)	8.0	1.43	4.2	1.03	42	6
F (15-19)	8.9	1.37	5.5	1.20	36	20
G (10-14)	7.1	1.68	4.7	1.28	27	53
H (5-9)	7.5	>3.00	3.8	1.68	2	65
I (>5)			2.2	1.82		8
Total	**9.5**	**1.46**	**4.3**	**1.46**	**16 7**	**153**

Researcher mobility

The advantage of using CVs is that they facilitate the study of 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 reported. We have in total 1 472 moves, but not all of these represent mobility from one institution to another, as we also record all changes in position (appointments). The mobility between institutions can be counted to a total of 594, which indicates that, on average, researchers have almost two inter-institutional moves during their career. Of course, as discussed above the distribution of mobility is uneven and some researchers are mobile while others are less mobile.

If we group all researchers in three: low, medium and high mobility we find that the most mobile researchers, a small group (18 persons) have the highest field citations score (=1.80), but the difference between low and medium is not that significant (low=1.42; medium=1.50) . But, we want to know more about the flows of people, how are they mobile? Table 7 shows the mobility patterns from one institution (row) to another institution (column). Although post doc experience is of interest in this context, we have decided not to include that information. Post doc mobility can be seen as a form of “institutionalized” mobility, valuable as such but with a specific logic. Each instance of mobility is registered as a change from row to column. There are 15 moves from UU (Uppsala University) to KI (Karolinska Institute). There are 50 moves from UU and only 31 researchers coming to 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 KI, i.e. a positive score (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 score (surplus) while Umeå University (UMU), Stockholm University (SU) and KTH (Royal Institute of Technology) have a negative score (deficit). The Institute Sector seems to be a neutral player in this field of research. It receives as many as it supplies, which indicates that it has a stable situation regarding financing of research.

Table 7. 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.

Searching for trends in mobility

Next step is to utilise the data for a study of traits of mobile persons and the direction of their mobility. We are dealing with the Swedish research landscape and the competition between universities to find competence and excellence. Some are winners and some are losers. Is it possible to put this into figures? In order to achieve this we propose a switch from the analysis in the last section, where we used moves between institutions to persons moving instead. Table 8 shows the result of this analysis. The table indicate the number of persons moving and the performance figure

(CPP/FCS) for these researchers. The table should be read in the following manner: A group of researchers start their career at KI and they are still there or have done a career move to other places. This group of 73 persons have a CPP/FCS of 1.49, i.e. 49 per cent above average. Uppsala (UU) loses a group of 15 researchers that now have appointments at KI. This group is considered as high performers with a CPP/FCS of 1.59. Uppsala has a large group that stays or comes back to UU and these "local heroes" are also very good performers (1.57).

The bottom row line gives the average CPP/FCS for all 45 persons who 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 rates makes it possible to draw some preliminary conclusions. Karolinska is the big winner in the competition for human capital in Swedish oncology research. Uppsala loses people with 1.50 and receives human capital with 1.30. The same applies to Umeå. Lund University has a positive exchange and this is also the case for Linköping University and Stockholm University. Figures for "local heroes" are often markedly lower than for the mobile group, except for Uppsala.

Table 8. Mobility and Citation Rate (CPP/FCS)

	→ KI	→ UU	→ LU	→ UMU	→ GU	→ SU	→ LIU	AVG CPP/FCS →
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/FCS	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

Finally, we will try to integrate cluster analysis and mobility analysis. Is there a typical path for the different cluster groups? There are 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. It should be noted that we are investigating mobility in relation to where researchers get their PhD and where they end up as professors. A first conclusion would be that mobility does not yield any significant differences between male and female investigators.

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 scientists in this group with a PhD from abroad. They tend to move around in Europe before finding 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 ends at the same university. This reflects the low mobility of researchers in Sweden.

Group 3 are the most prestigious researchers, characterised by their excellence, seniority, and "watchdog" experience. Starting their career at Karolinska or Uppsala, also this group tend to end up at their mother university. A short time period between PhD and professorship does not seem to be connected to an overall pattern of mobility as this group, after a short career tour, often are recruited back to their home university.

Lastly, we have the mobility of the entrepreneurial group. Here we should notice that post doc experience differentiates between groups. In general, about half of each group have done a post doc abroad, mainly in the United States. The entrepreneurial group deviate from this quite strong rule: to a larger extent, about 75 per cent, they have been abroad during their post doc. They tend to stay longer, but when they come back to Sweden they find themselves a position at the best institutions, i.e. Uppsala University and Karolinska Institute.

In conclusion, as soon as we want to use large numbers we always end up with the large group of "local heroes" – researchers who receive their PhD at one university and end up as professors at the same university. And this overall pattern does not seem to be gender specific.

Discussion

The main contribution of this paper is to include citation analysis and test it on a fairly large dataset of CVs. Using field normalized citations and papers during a specific period of time have given new insights and new perspectives on the problems of gender, mobility, collaboration and productivity. A first result is that gender is related to productivity in a double way: female scientists may be less productive in both normal and fractional paper counts, but their impact in field normalized citations is better than that of their male colleagues. This is the paradox of gendered productivity patterns that seems to be relevant in many areas of research.

By applying cluster analysis we have been able to pinpoint some of the patterns underlying medical research in Sweden. The four groups of researchers – Mobile, Immobile, Excellence and Entrepreneur – reveal important differences when it comes to variables like gender and impact of research. Female scientists are over-represented in the first two groups and underrepresented in the latter groups. One implication 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 lie behind and can explain why this is the case. The female tendency to lower publication activity but higher citation rates makes it necessary to find ways of translating their revealed excellence into strategies for resource assembling. This they will have to do in relation to research foundations and research councils that, up to now, have been predominantly interested in numbers of publications (Brouns, 2000; Sandström & Hällsten, 2008). In the pessimistic view, female researchers will not be able to assemble resources for their research. Instead, they will have to align with older and maybe not so innovative male colleagues who control the funding resources and still are the main recipients of significant funding resources.

Collaborative research is productive but less collaborative research is cited relatively more frequently. This result goes against many of the propositions of earlier research. Lee & Bozeman (2005) hold the door open for conclusions like these as they do not have a quality indicator, and the conclusion here is that these fundamental questions should be further investigated. Mobility between Swedish universities seems to be a process of selecting out good researchers from less productive researchers. Therefore, mobility can be understood as determined to a large extent by researcher impact (quality), or, at least, as a mixture of mobility (self selection) and reputation.

References

Allison, PD & Long, JS 1987 Interuniversity mobility of academic scientists. American Sociological Review 52(Oct):643–652.

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

Canibano, C 2008 Measuring and assessing researcher mobility from CV analysis: the case of the Ramon y Cajal programme in Spain. Research Evaluation 17 (1): 17-31.

Cesaroni, F & Gambardella, A 2003 Research productivity and the allocation of resources in publicly funded research programmes. (Eds.) Geuna et al. Science and Innovation: rethinking the rationales for funding and governance. Cheltenham, UK: Edward Elgar, pp.202–232.

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.
Lee, S & Bozeman, B 2005 The Impact of Research Collaboration on Scientific Productivity. Social Studies of Science 35 (5):673–702.
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.
Merton, RK 1973 The Sociology of Science: theoretical and empirical investigations. Chicago: Chicago University Press.
Narin, F & Hamilton, KS 1996 Bibliometric Performance Measures. Scientometrics 36 (3): 293–310.
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.
Prpic, K 2002 Gender and Productivity Differentials in Science. Scientometrics 55 (1): 27-58.
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.
Sandström, U & Hällsten, M (forthcoming) Research Quality and Diversity of Funding: A Model for Relating Research Money to Output of Research, Scientometrics
Sandström, U & Hällsten, M 2008 Persistent Nepotism in Peer Review. Scientometrics 74 (2): 175–189.

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

Docent and associate professor Ulf Sandström is with the Dept of Social Change and Culture at Linköping University, 581 83 LINKÖPING, SWEDEN. Email: ulfsa@isak.liu.se. Tel: +46708-137376. Support from FORMAS for the research project 151-2007-159 is acknowledged. The author would like to express his gratitude to Ulf Heyman, Martin Hällsten, Erik Sandström, and one anonymous referee for their valuable remarks.