

Bibliometrics as a research assessment tool - impact beyond the impact factor

Jonas Lundberg
Medical Management Centre
Karolinska Institutet



**Karolinska
Institutet**

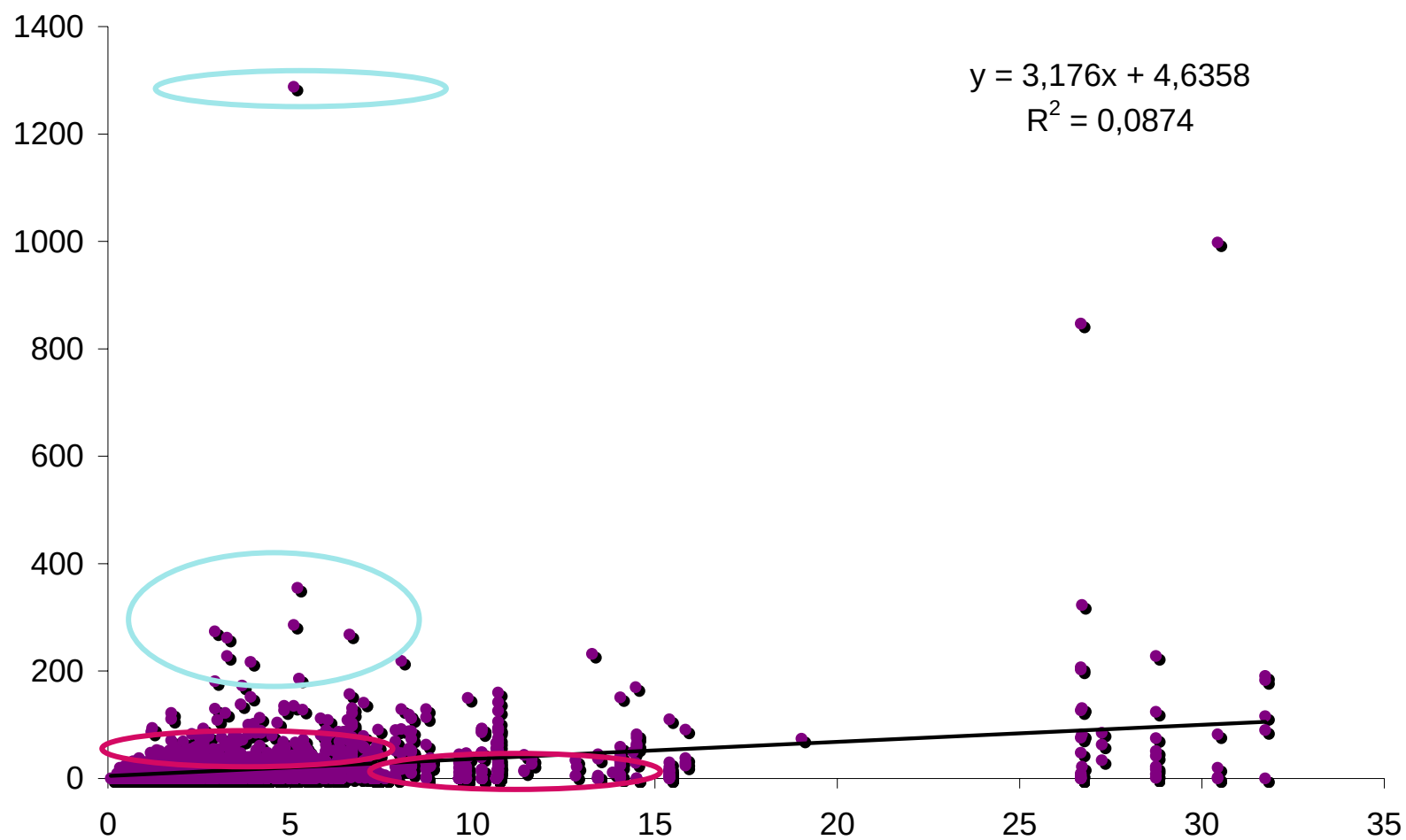
Bibliometrics – *“the application of mathematical and statistical methods to books and other media of communication”*

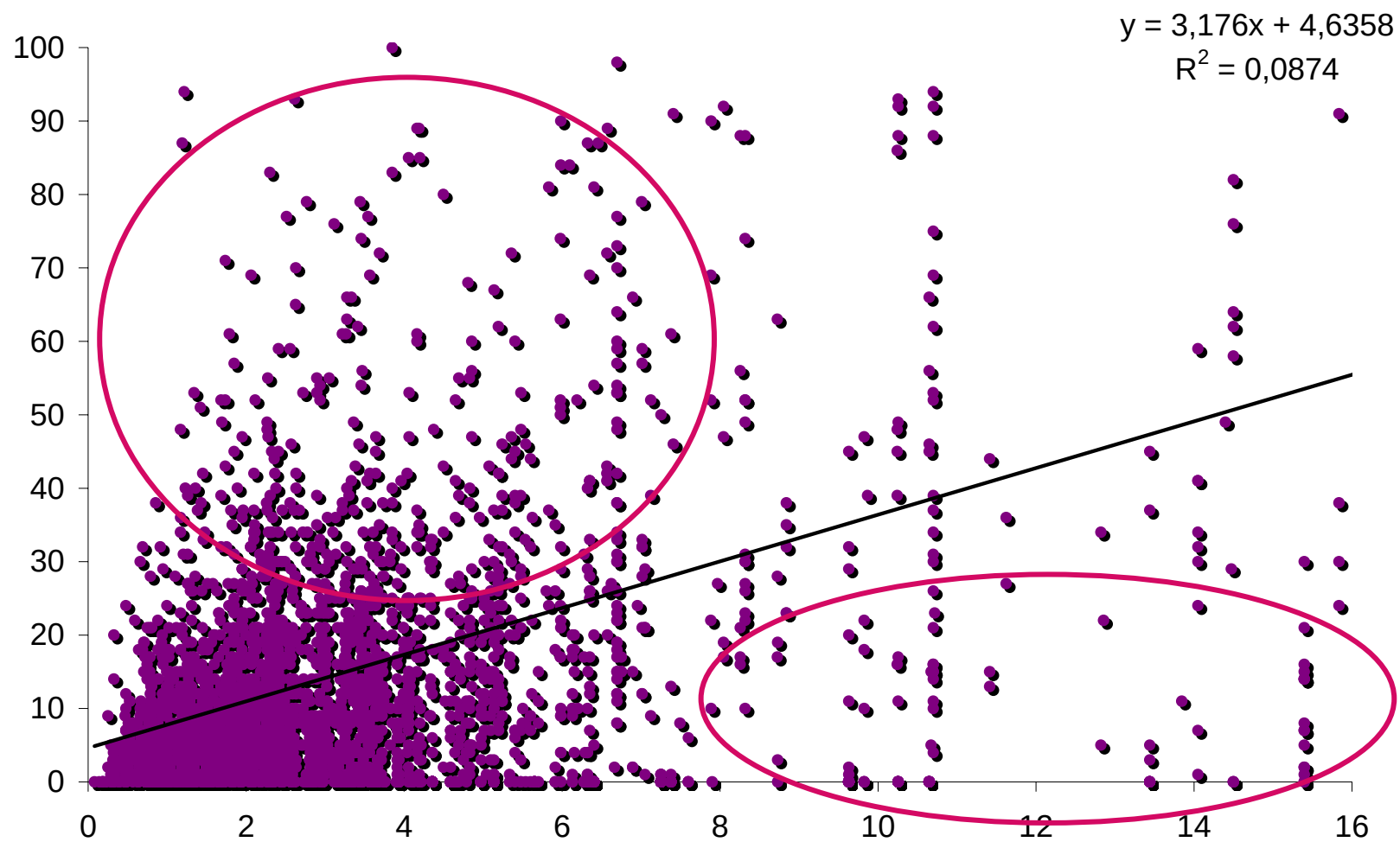


**Karolinska
Institutet**

Impact Factors vs Individual citations

Example: KI publications 1997





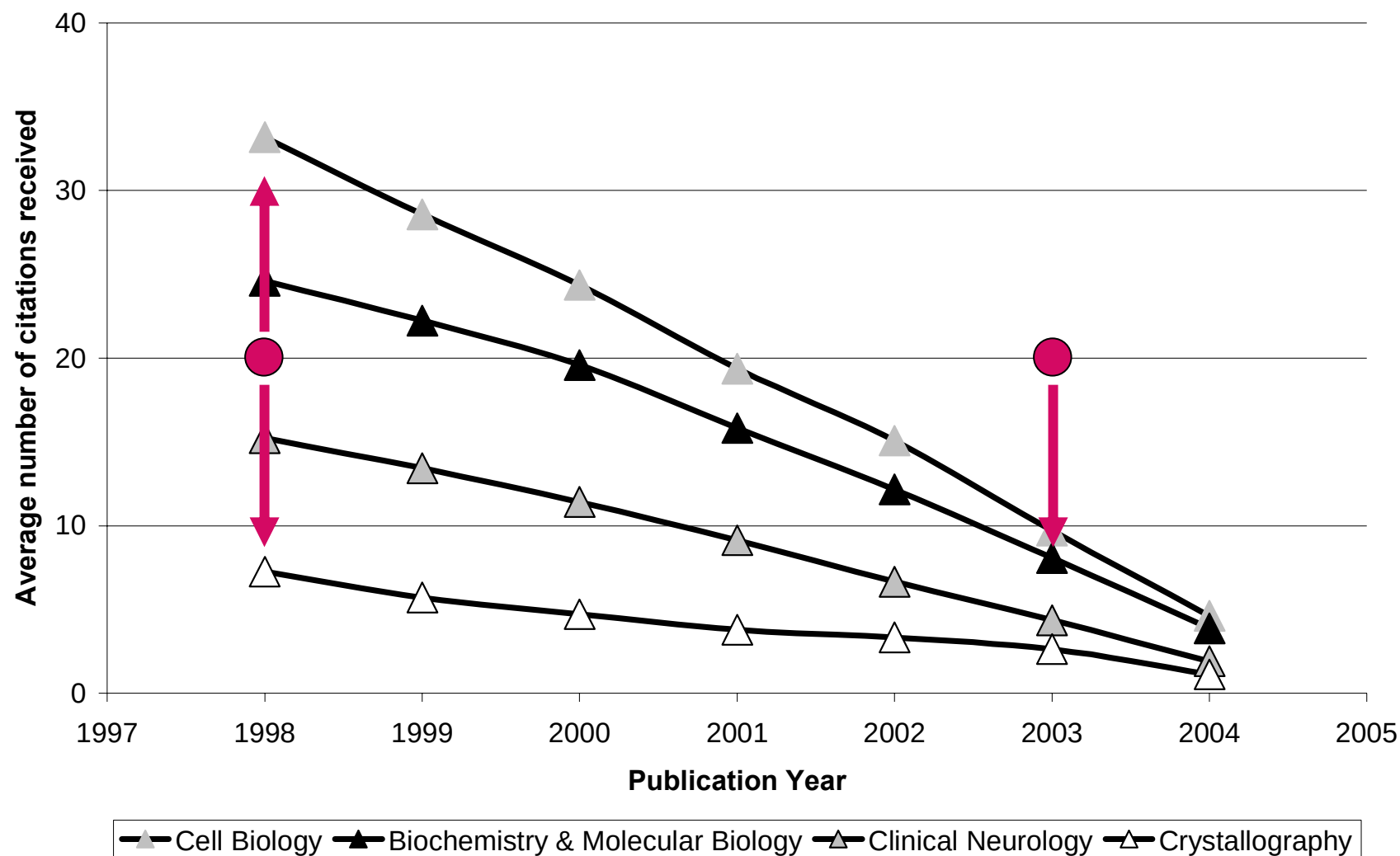
n	Journal	IF
46	New England Journal of Medicine	44.0
35	Circulation	11.6
30	The Lancet	23.4
17	Radiology	5.4
16	JAMA	23.3
12	International Journal of Oral & Maxillofacial Implants	1.4
11	Clinical Oral Implants Research	1.9
9	Pediatrics	4.3
8	American Journal of Cardiology	3.1
8	British Journal of Surgery	3.7
8	Clinical Therapeutics	3.0
8	European Journal of Vascular and Endovascular Surgery	2.0

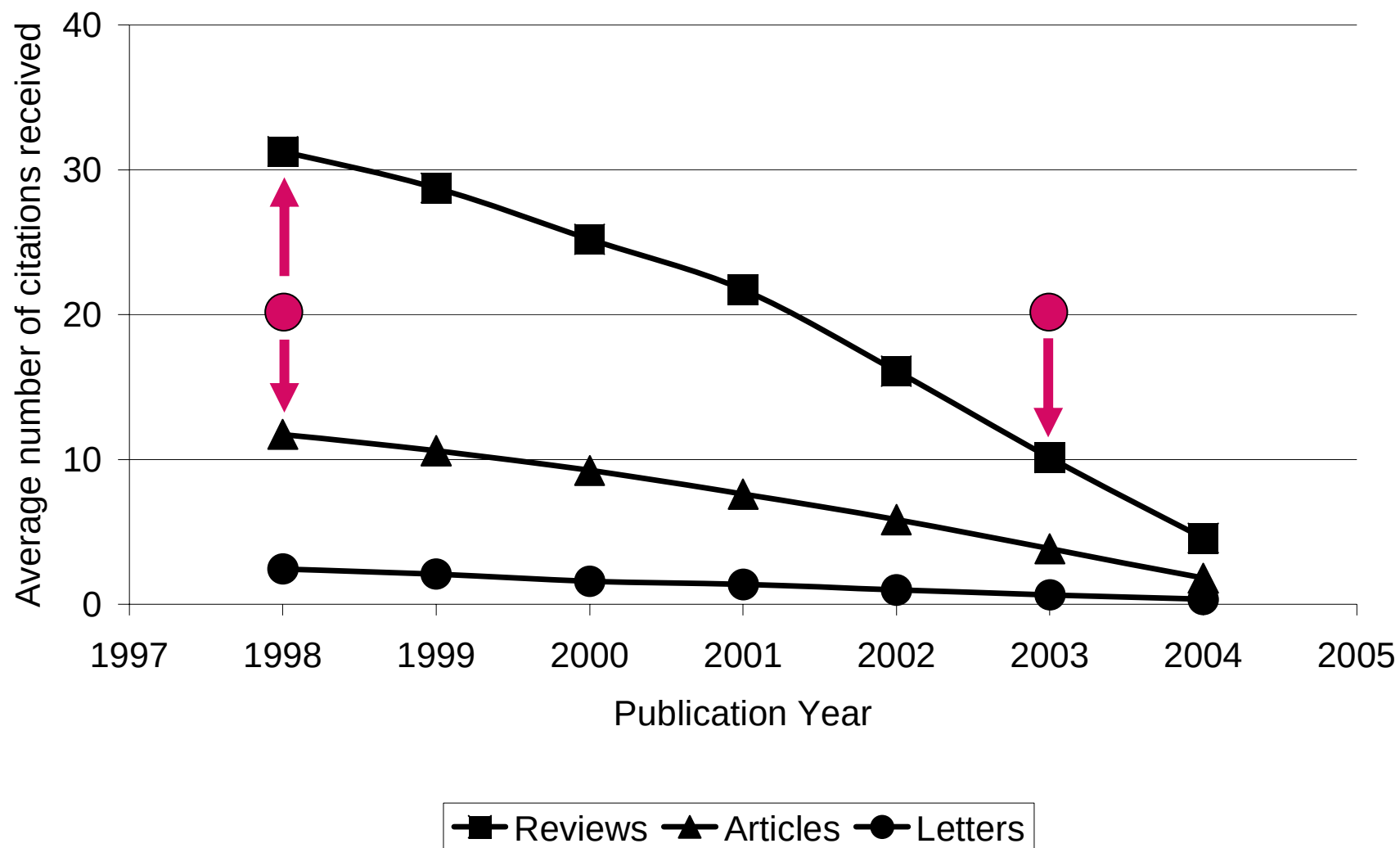


**Karolinska
Institutet**

A publication has been cited twenty times. That is...

- a) Good**
- b) Bad**
- c) Don't know**





$\overline{c}_f$

item oriented field normalized citation score average

- A department has three publications, X, Y och Z.
 - X: 30 citations (Normal article, 1999, Oncology)
 - Y: 60 citations (Review, 1997, Biochemistry)
 - Z: 40 citations (Note, 2001, Nursing)
- World average
 - Normal article, 1999, Oncology: 20
 - Review, 1997, Biochemistry: 60
 - Note, 2001, Nursing: 20

$$\begin{aligned} & \text{■ } c/\mu_f \\ & \quad \rightarrow X: 30 / 20 = \mathbf{1,5} \\ & \quad \rightarrow Y: 60 / 60 = \mathbf{1} \\ & \quad \rightarrow Z: 40 / 20 = \mathbf{2} \\ & \text{■ } \overline{c}_f \\ & \quad \rightarrow (1,5+1+2)/3 = \mathbf{1,5} \end{aligned}$$

$\bar{c}_f$

item oriented field normalized citation score average

Article	Type	Year	Journal	Field	c	μ_f	c_f
A	Article	2003	J Appl Crystallogr	Crystallography	12	2.6	4.6
B	Article	2003	J Cryst Growth	Crystallography	5	2.6	1.9
C	Review	2000	Nat Rev Immunol	Immunology	66	33.4	2.0
D	Article	2000	J Immunol	Immunology	17	17.0	1.0
E	Letter	2001	Am J Hum Genet	Genetics & Heredity	17	9.7	1.8

$$\frac{1}{P} \sum_{i=1}^P \frac{c_i}{[\mu_f]_i} = \frac{4.6 + 1.9 + 2.0 + 1.0 + 1.8}{5} \approx 2.2$$

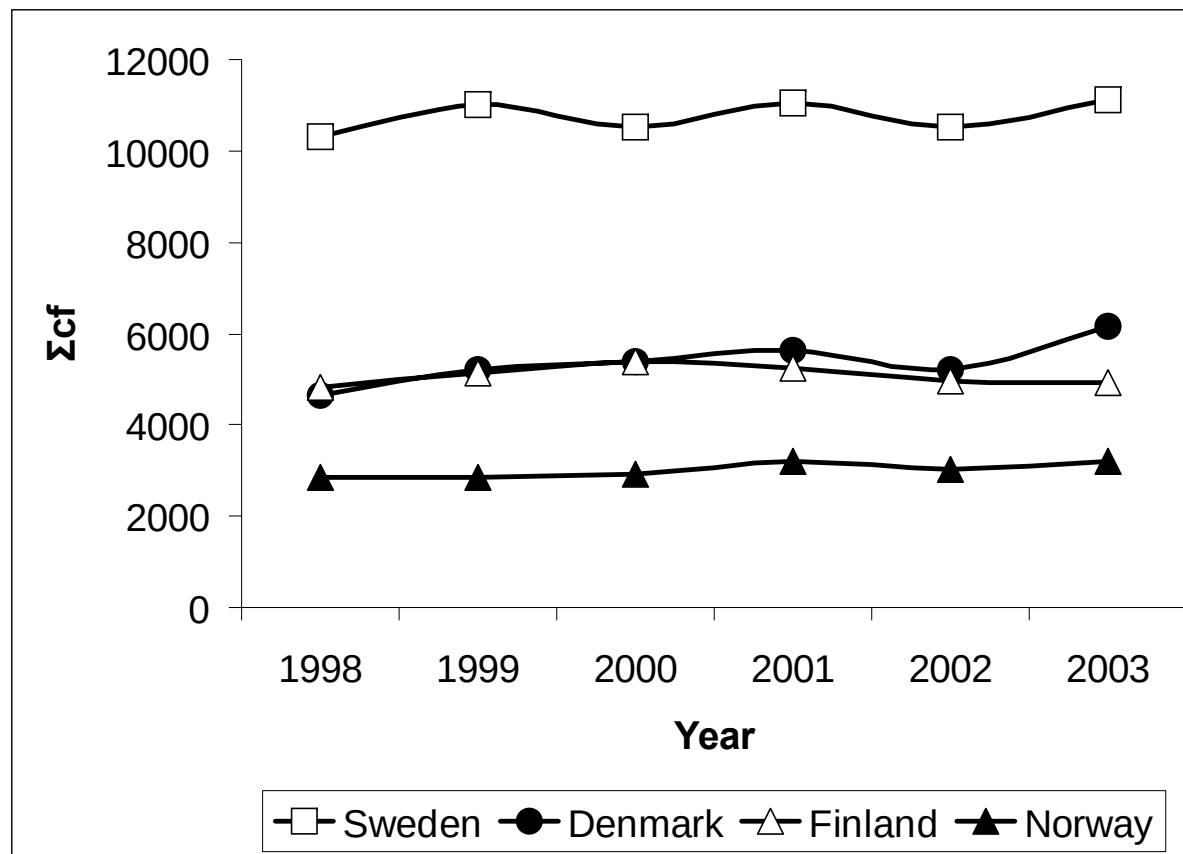
$[\bar{c}]_f$

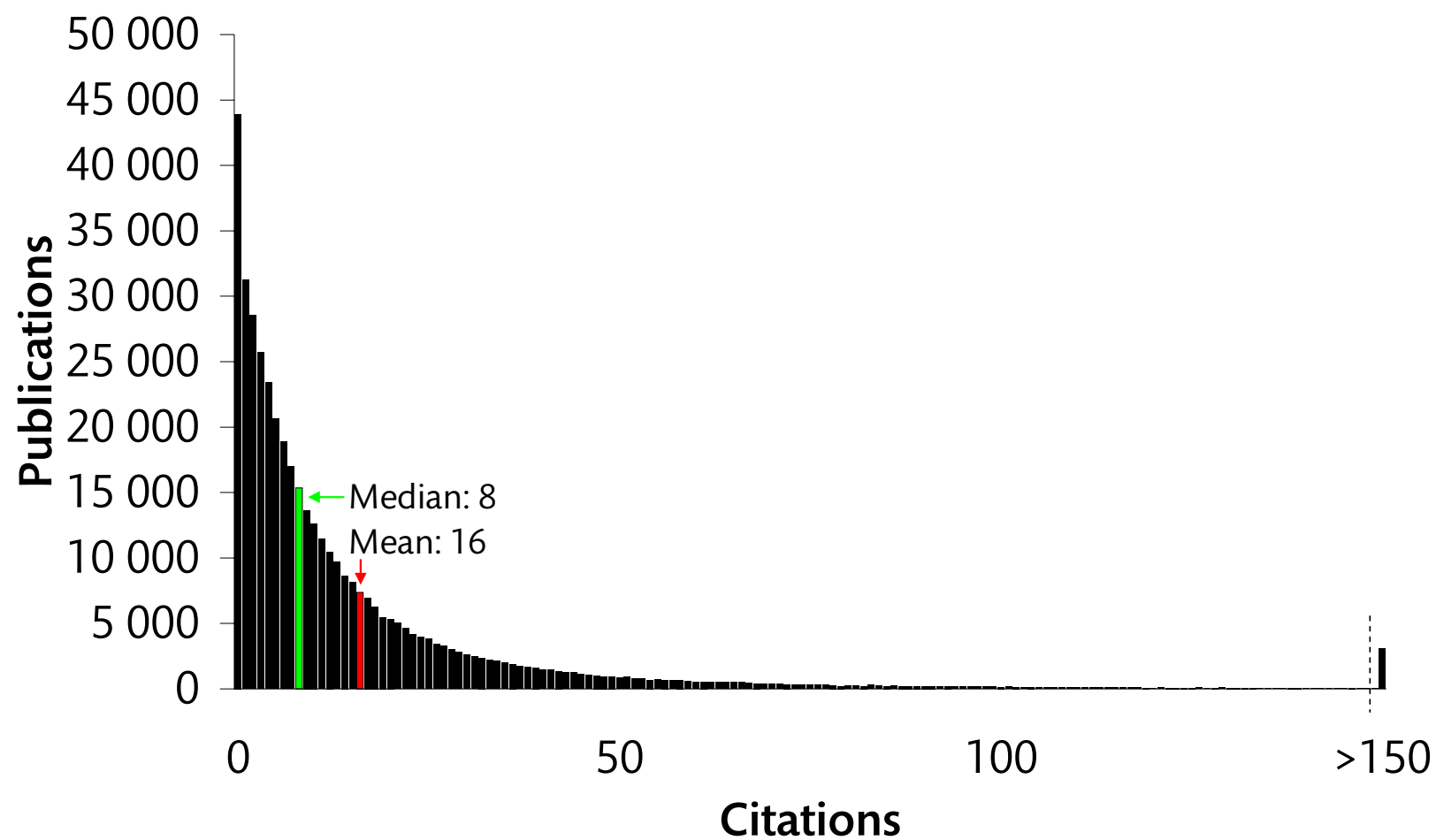
The crown indicator

Article	Type	Year	Journal	Field	c	μ_f
A	Article	2003	J Appl Crystallogr	Crystallography	12	2.6
B	Article	2003	J Cryst Growth	Crystallography	5	2.6
C	Review	2000	Nat Rev Immunol	Immunology	66	33.4
D	Article	2000	J Immunol	Immunology	17	17.0
E	Letter	2001	Am J Hum Genet	Genetics & Heredity	17	9.7

$$\frac{\sum_{i=1}^P c_i}{\sum_{i=1}^P [\mu_f]_i} = \frac{12 + 5 + 66 + 17 + 17}{2.6 + 2.6 + 33.4 + 17.0 + 9.7} \approx \frac{117}{65.3} \approx 1.8$$

ΣC_f total field normalized citation score





$\overline{c}_{fz[ln]}$

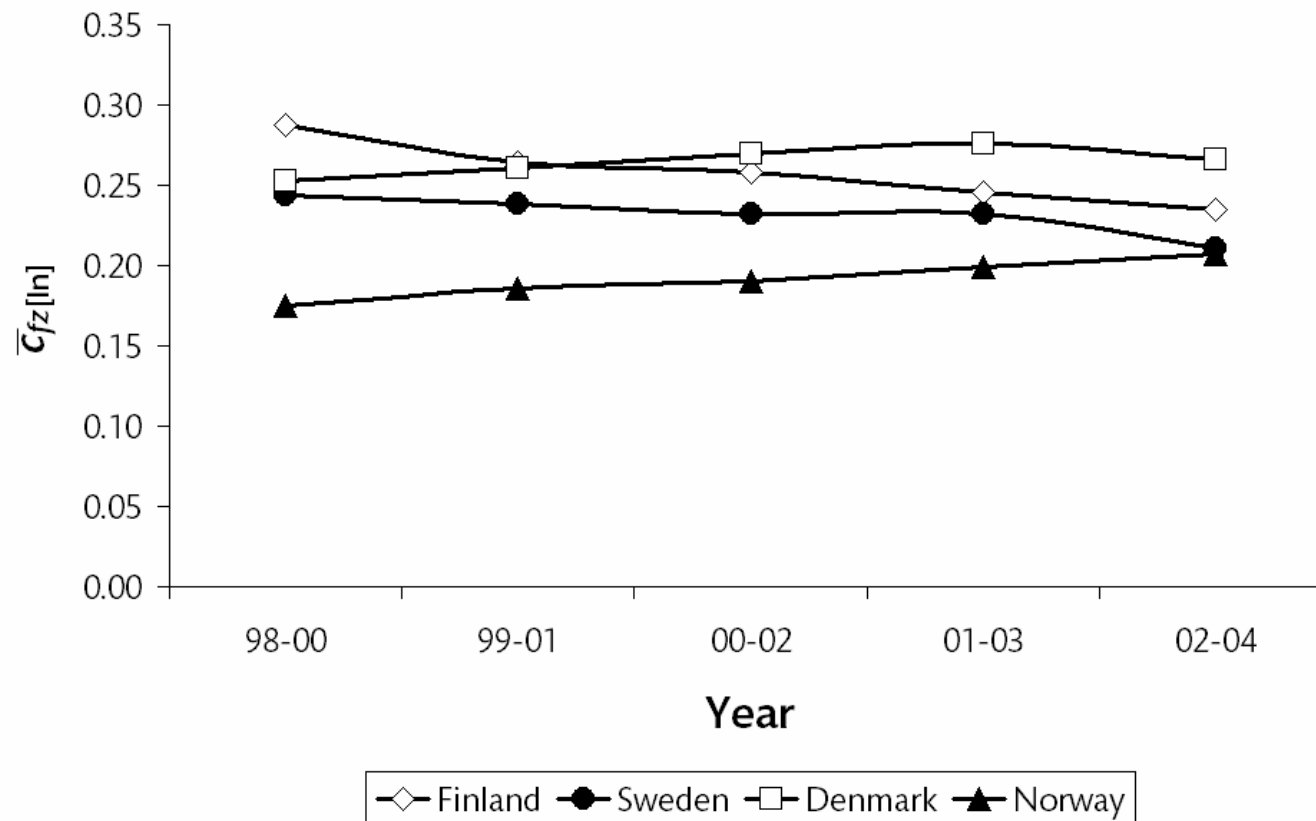
item oriented field normalized logarithm-based citation
z-score average – “citation z-score”

Article	Field	c	μ_f	σ_f	c_f	$\ln(c+1)$	$\mu_{fz[ln]}$	$\sigma_{fz[ln]}$	$c_{fz[ln]}$
A	Crystallography	12	2.6	22.1	4.6	2.6	0.9	0.8	2.2
B	Crystallography	5	2.6	22.1	1.9	1.8	0.9	0.8	1.2
C	Immunology	66	33.4	76.1	2.0	4.2	2.7	1.3	1.1
D	Immunology	17	17.0	27.4	1.0	2.9	2.3	1.2	0.6
E	Genetics & Heredity	17	9.7	32.2	1.8	2.9	1.6	1.2	1.1

$$\frac{1}{P} \sum_{i=1}^P \frac{\ln(c_i + 1) - [\mu_{f[ln]}]_i}{[\sigma_{f[ln]}]_i} = \frac{2.2 + 1.2 + 1.1 + 0.6 + 1.1}{5} \approx 1.2$$

$\bar{c}_{fz[ln]}$

item oriented field normalized logarithm-based citation
z-score average – “citation z-score”

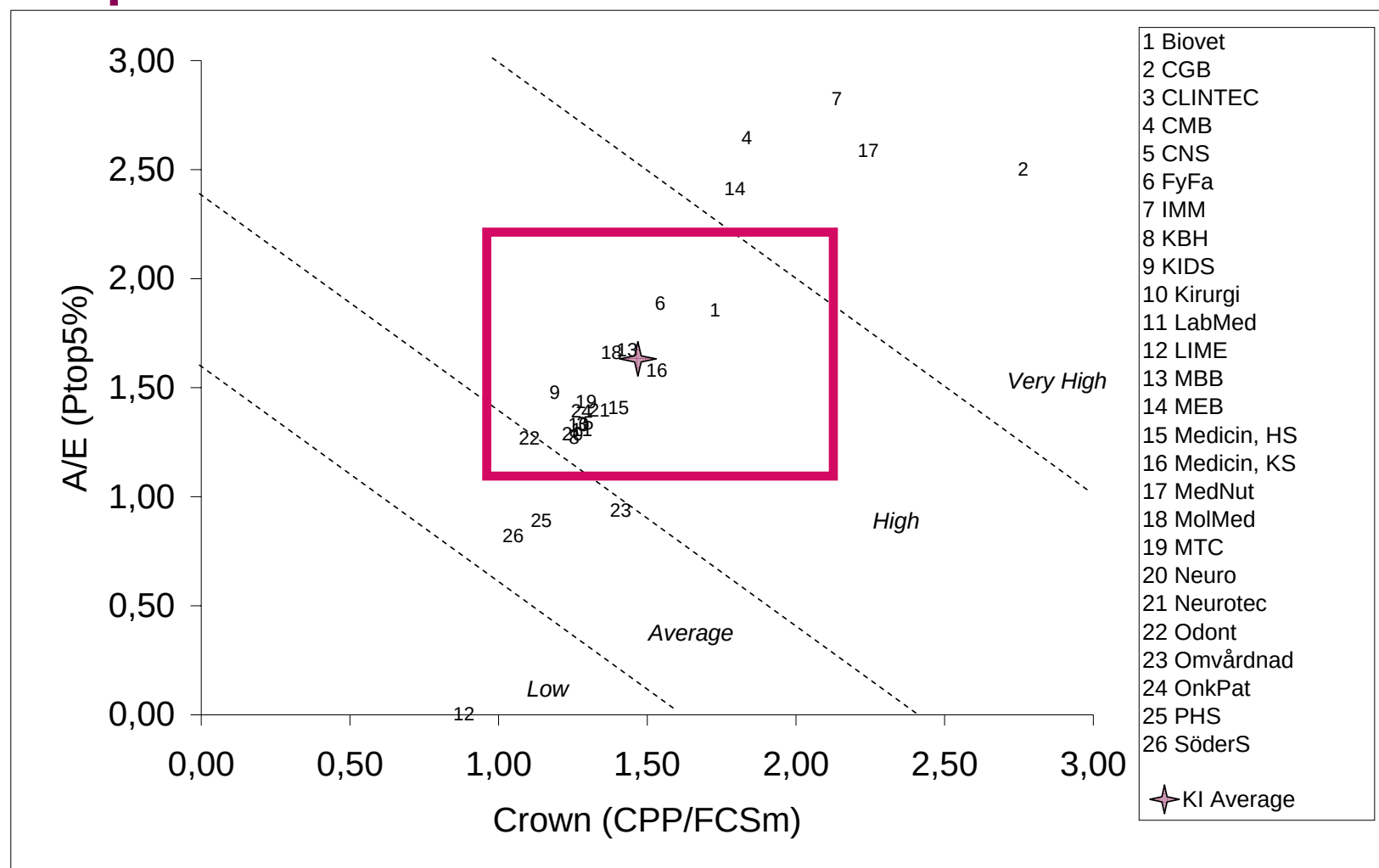


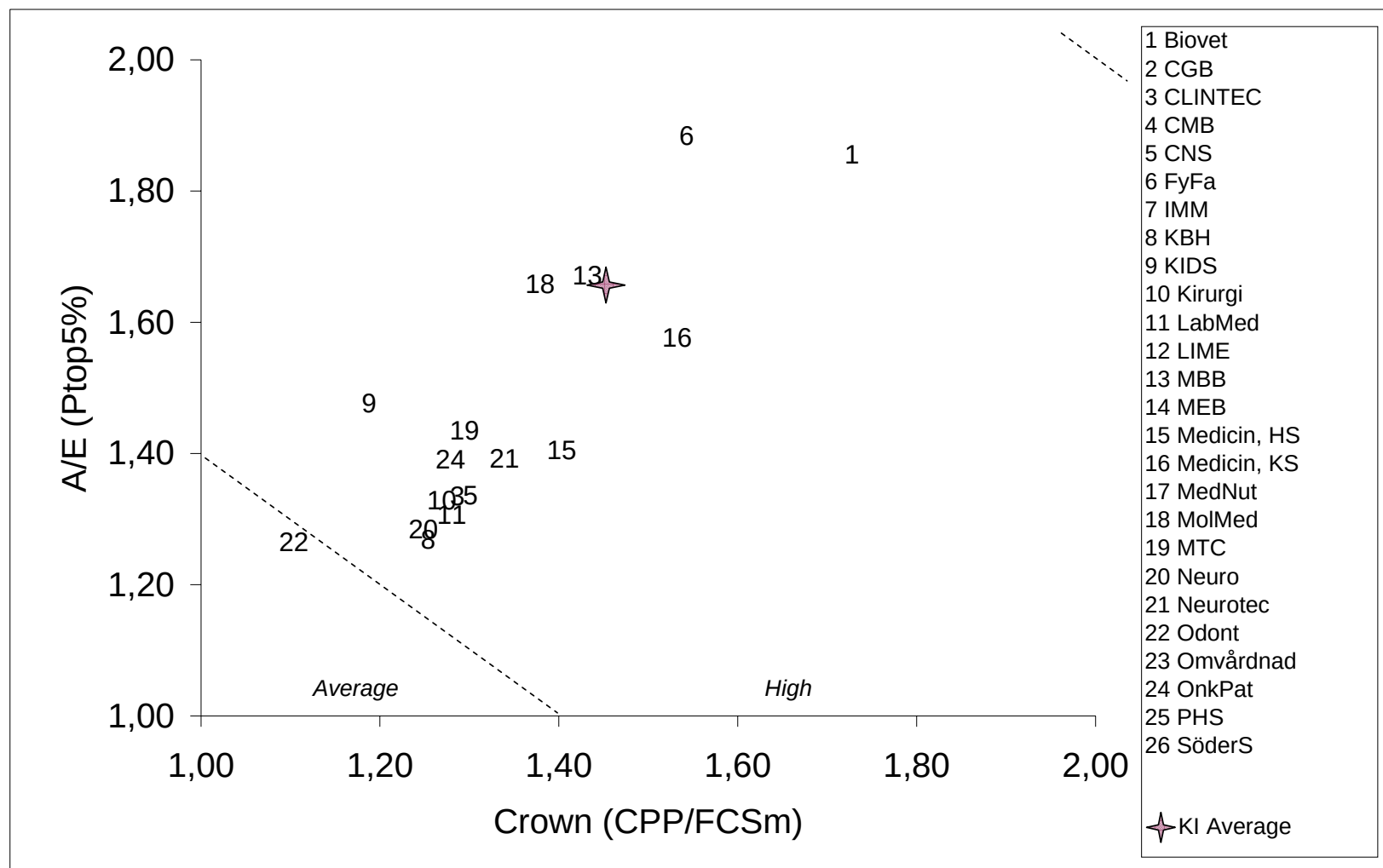
Share top 5%, example

- A department has three publications, X, Y och Z.
 - X: 30 citations (Normal article, 1999, Oncology)
 - Y: 60 citations (Review, 1997, Biochemistry)
 - Z: 40 citations (Note, 2001, Nursing)
- To be among the the top 5% most highly cited
 - Normal article, 1999, Oncology: 60
 - Review, 1997, Biochemistry: 103
 - Note, 2001, Nursing: 36

- Is the publication among the 5% most highly cited?
 - X: $30 < 60$ No
 - Y: $60 < 103$ No
 - **Z: $40 > 36$ Yes**
- Share of publications among top 5%:
 - **$1/3 = 33\%$**
 - *(Indicator value: $33\%/5\%=6,7$)*

Departments

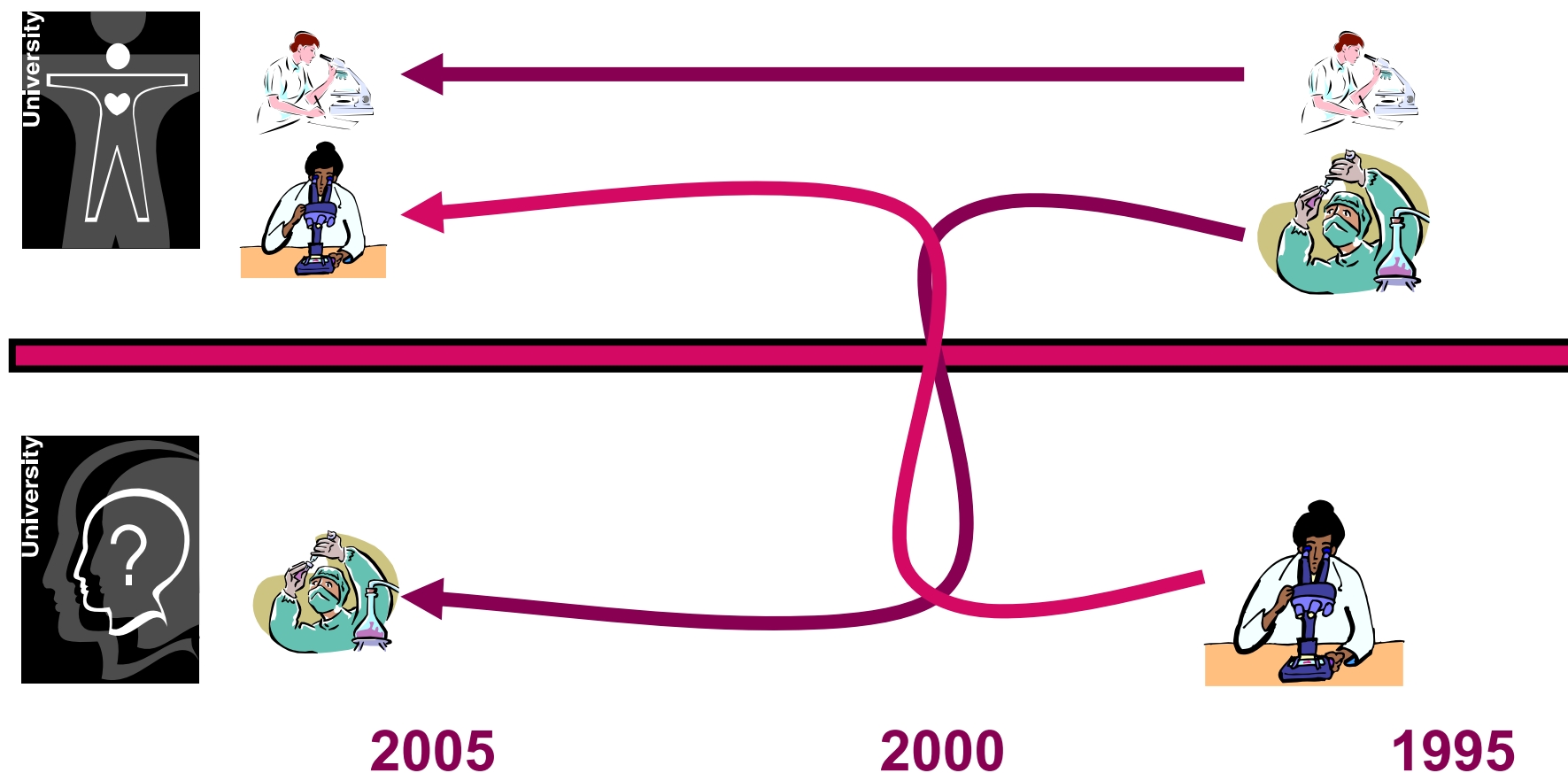




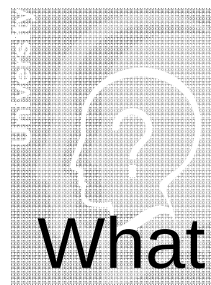
Decisions to make

- All bibliometric assessments
 - level of aggregation
 - time frame
 - retrospective or prospective focus
 - counting scheme
 - citation window
 - document types
- Additional for resource allocation
 - What indicators to include use
 - Weights

Retrospective vs prospective bibliometrics



Retrospective vs prospective bibliometrics



Retrospective bibliometrics

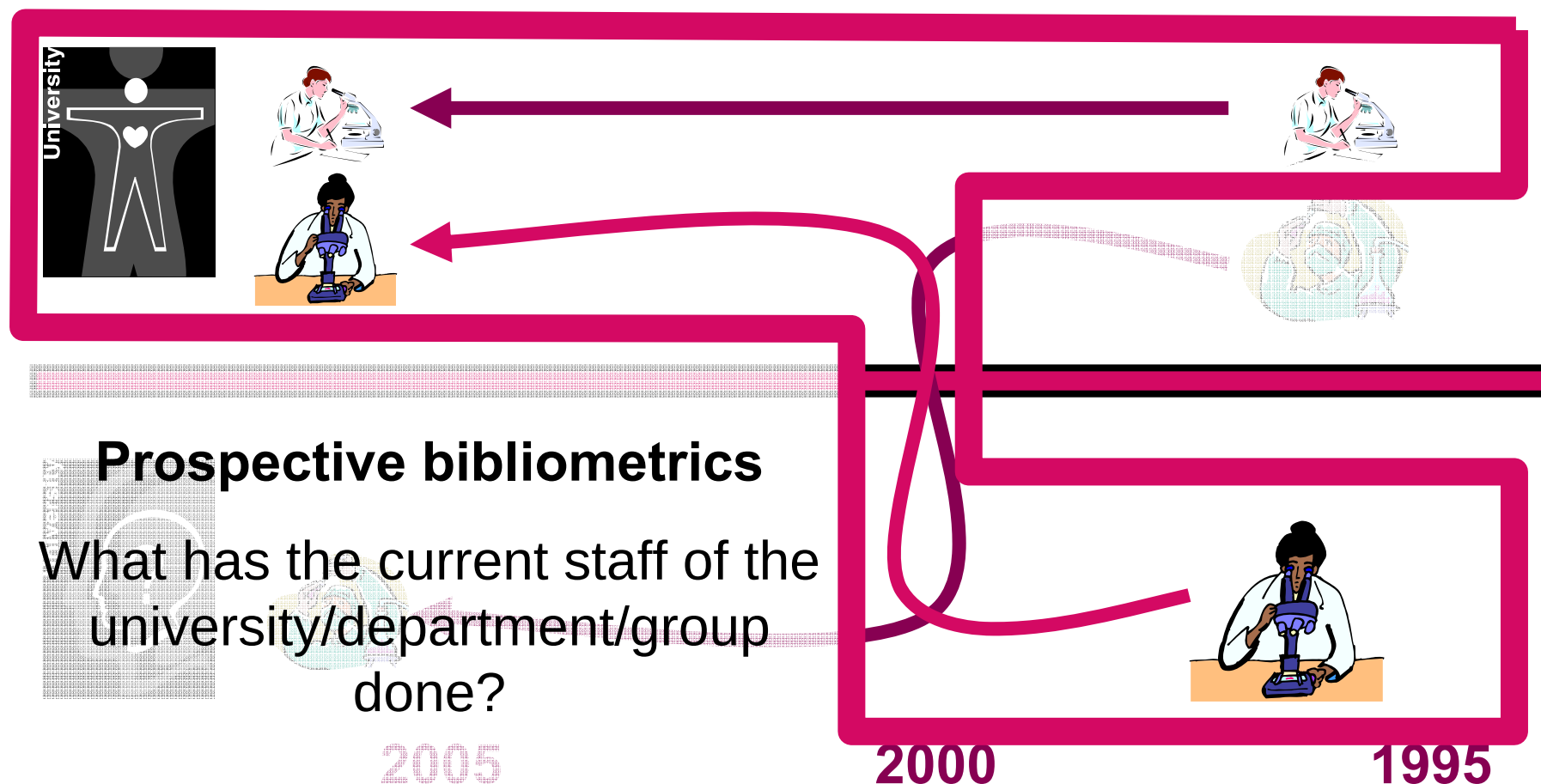
What has the university/department/group produced?

2005

2000

1995

Retrospective vs prospective bibliometrics



Common counting schemes

- *Fractional counting* – if n units (authors, institutions, countries, etc.) have contributed to the paper, each takes the value $1/n$ for the paper.
- *First address count* – a paper is assigned to one unit only, on the basis of the first address in the address list of a paper as included in the database.
- *Whole counting* (also called *full* or *integer counting*) – assigns a publication fully to each contributing unit.



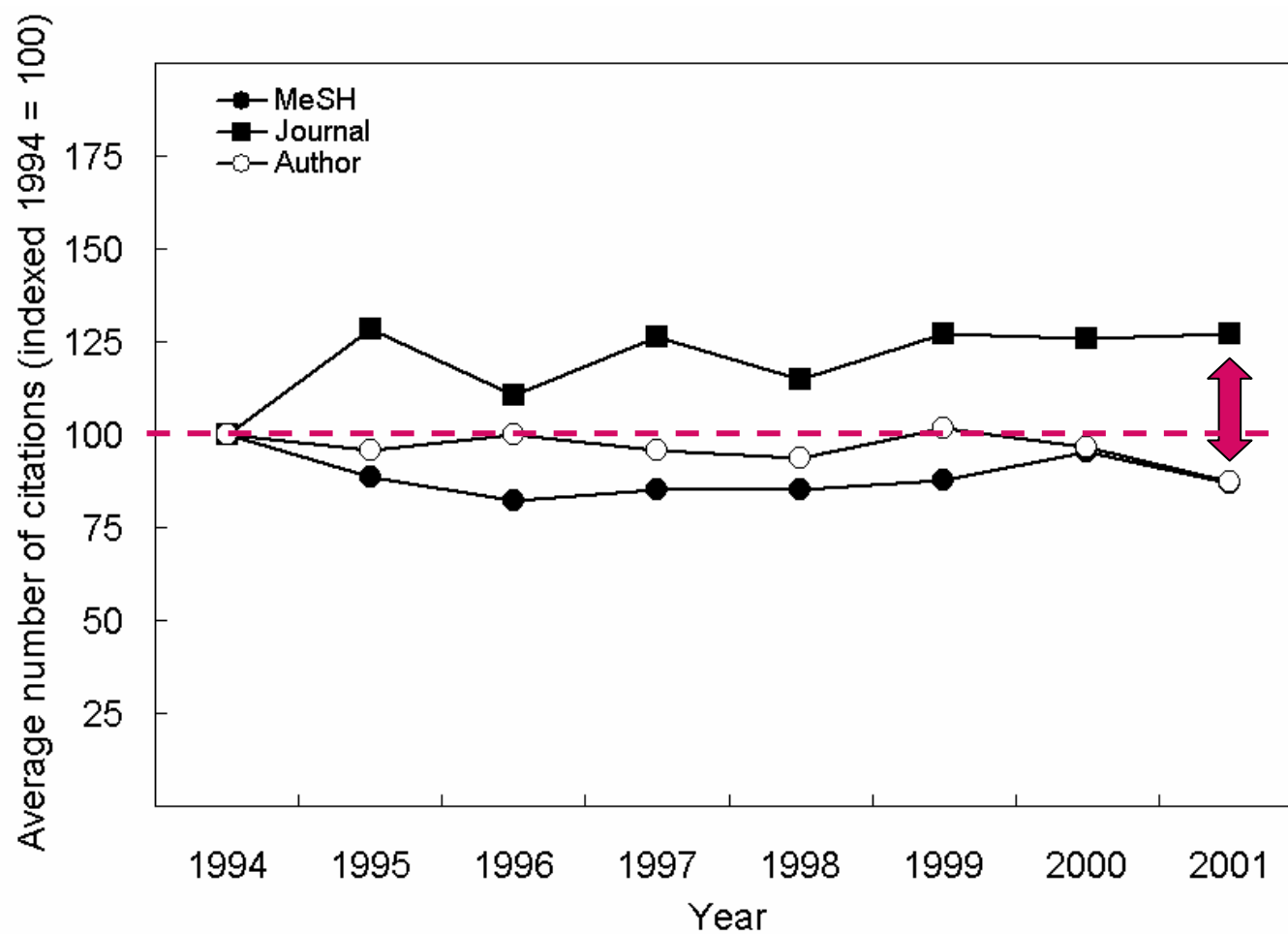
**Karolinska
Institutet**

Methodological considerations

Definition of research areas

Journal		
1	Journal of biological chemistry	Biochemistry & Molecular biology
2	International journal of cancer	Oncology
3	Bone marrow transplantation	Haematology Immunology Oncology Transplantation
4	PNAS	Multidisciplinary sciences
5	FEBS Letters	Biochemistry & Molecular biology Biophysics Cell biology
6	Blood	Haematology
7	Carcinogenesis	Oncology
8	Biochemical and biophysical research communications	Biochemistry & Molecular biology Biophysics
9	Cancer research	Oncology
10	Acta oncologica	Oncology

Is it better or just the same?



Funding & Co-authorship as collaboration indicators

	Funding (+)	Funding (-)	Total
Co-authorship (+)	64	341	405
Co-authorship (-)	31	0	31
Total	95	341	436



**Karolinska
Institutet**

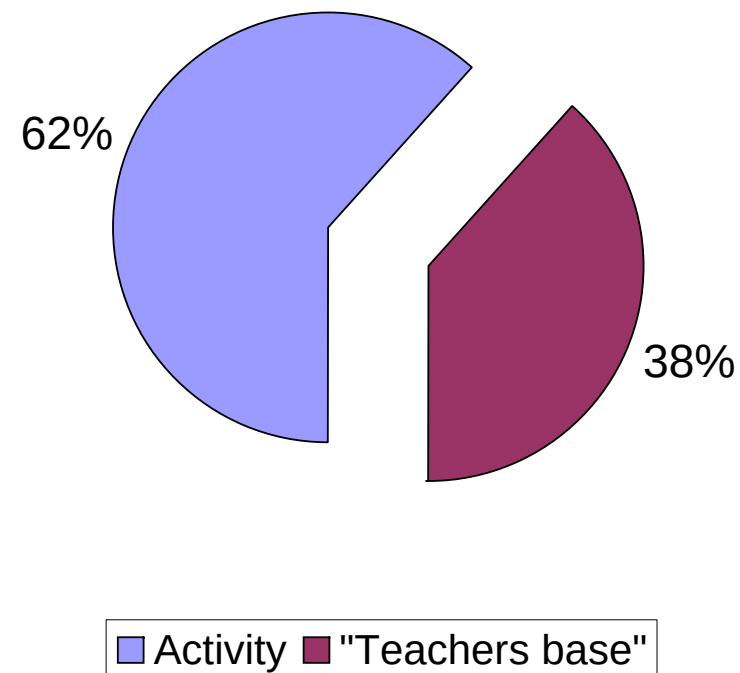
Bibliometrics and resource allocation



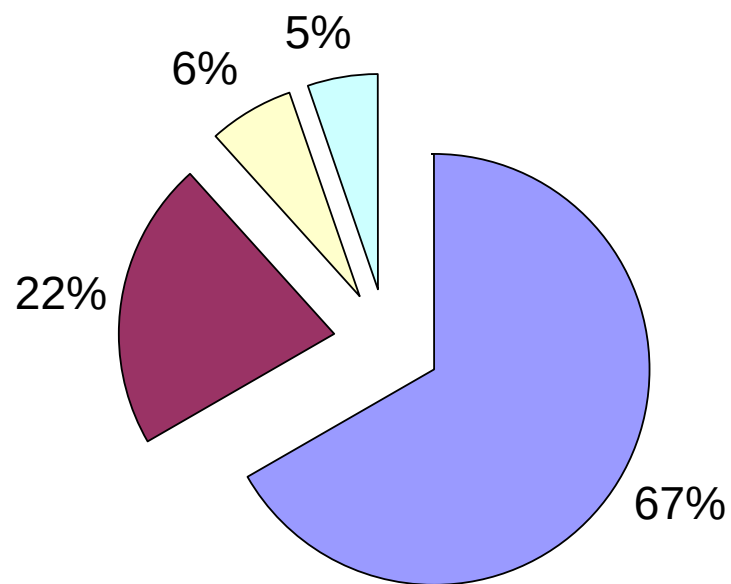
**Karolinska
Institutet**

Resource allocation system at Karolinska Institutet

Resource allocation



"Activity"



Example of a resource allocation model based on bibliometric data

A = Share of publications in journals with IF > 11 (2000 – 2003)

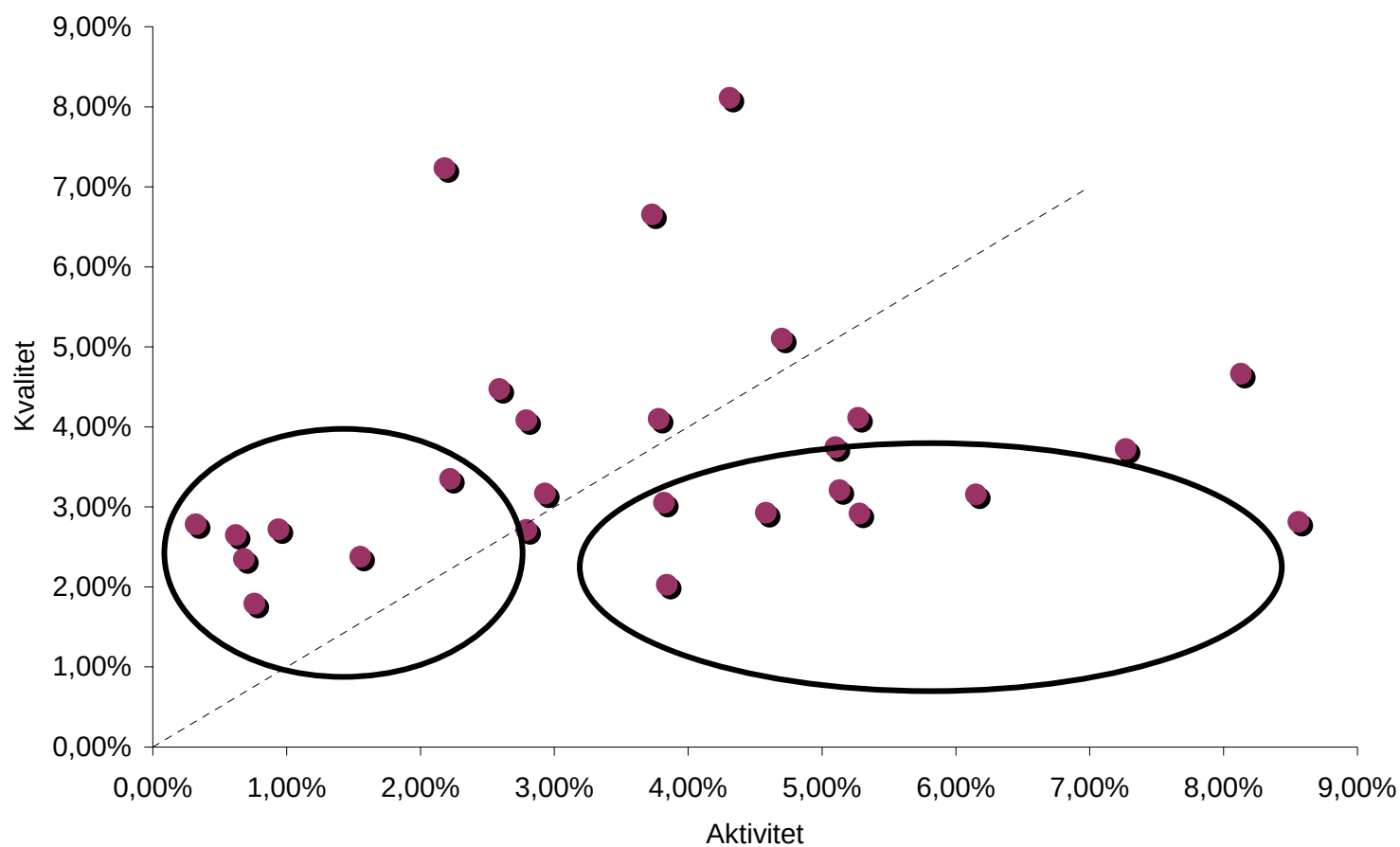
B = Field normalized citation score (2000 – 2003)

C = Share of publications among top 5% highly cited within research area (2000 – 2003)

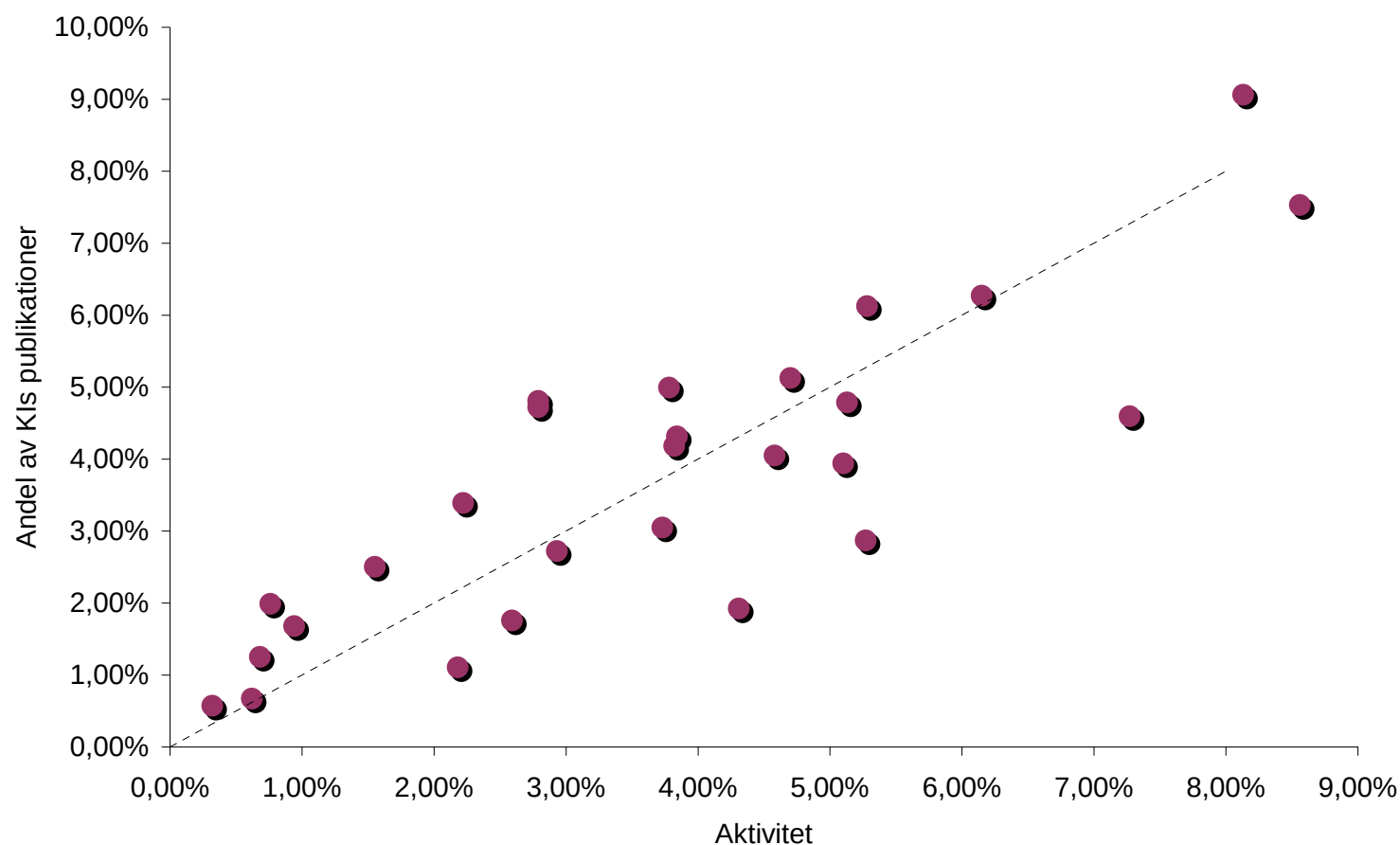
D = Publication count (2000 – 2003)

$$\left(\frac{A_i}{\sum_{i=1}^n A_i} \times 25\% + \frac{B_i}{\sum_{i=1}^n B_i} \times 50\% + \frac{C_i}{\sum_{i=1}^n C_i} \times 25\% \right) \times \frac{D_i}{\sum_{i=1}^n D_i}$$

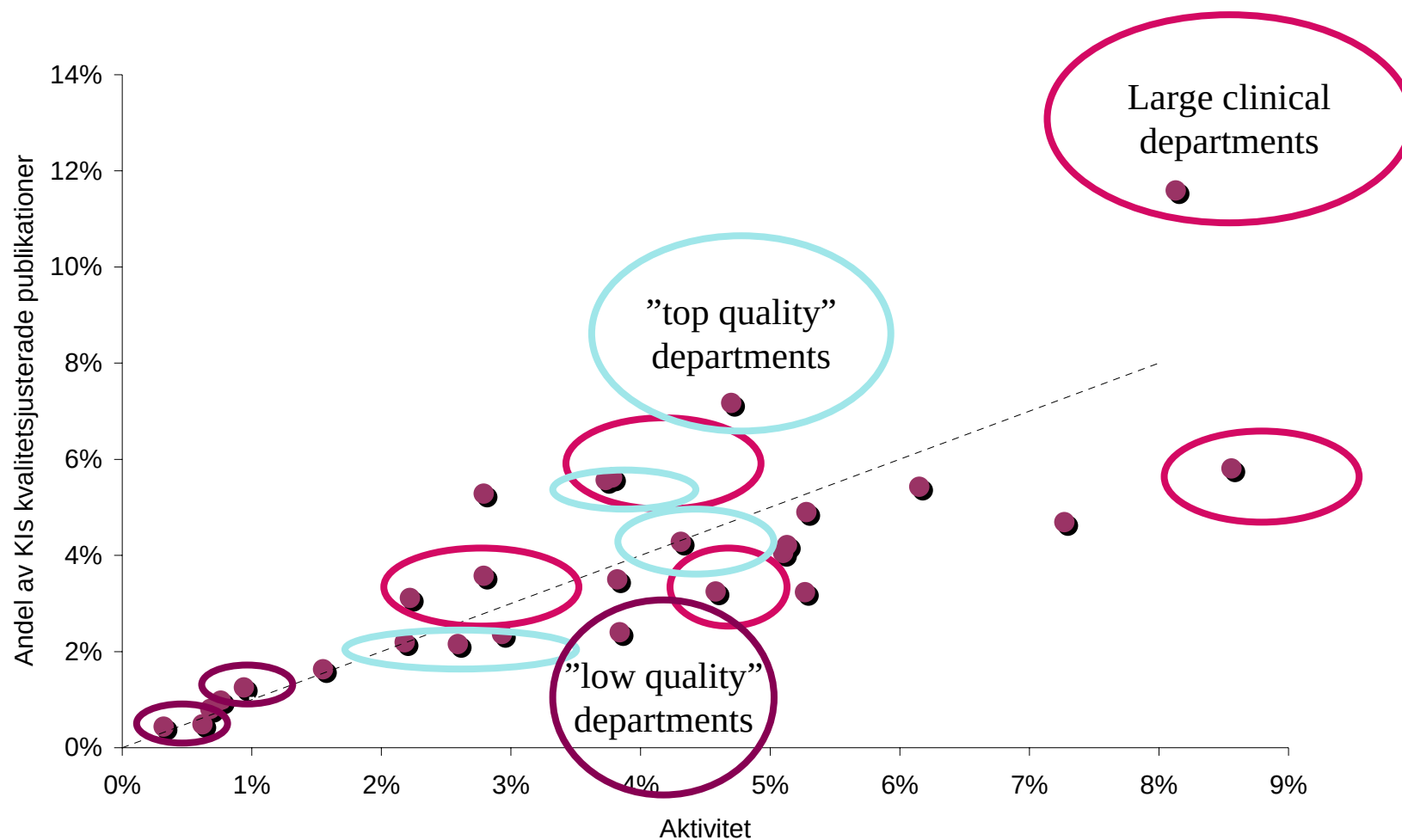
Average "bibliometric quality" (00-03) and activity (2005)



Number of publications (00-03) and activity (2005)



Quality adjusted publication score (00-03) and activity (2005)



Thank you!

Jonas Lundberg

e-mail: jonas.lundberg@ki.se

tel: +46 8 524 866 89

ki.se/bibliometrics