Bibliometric indicators

– definitions and usage at Karolinska Institutet

Catharina Rehn, Ulf Kronman and Daniel Wadskog, Karolinska Institutet University Library

Version 1.0

2007-08-22

This appendix to the *Bibliometric handbook for Karolinska Institutet* lists the most important bibliometric indicators together with their definitions, some comments on advantages and shortcomings of the different indicators, and their usage at Karolinska Institutet.

It may be noted that this list of indicators still is a work in progress. Some of the indicators are not fully described and some lack formulas and references. New updated versions of the indicator list are planned be released as the work with bibliometrics at Karolinska Institutet progresses.

First, some general notes on the definitions and the calculation of indicators in the appendix:

Inclusion or exclusion of self citations – see the handbook for more information – might affect the resulting indicator values, but not how the indicators are calculated. Self citations are therefore noted as a separate indicator, but not in the context of any of the other indicators. At Karolinska Institutet, we do not presently remove self citations while calculating our indicator values.

Fractionalization or any other form of weighting of publications between the contributing authors – see the handbook for more information – will affect most indicators. It will, however, not affect the basic calculation principles, and, for reasons of clarity, this aspect has been left out in the indicator descriptions. At Karolinska Institutet, we do not currently use any fractionalization or weighting while calculating our indicator values.

The validity of several of the indicators improves if the authors themselves validate or supply information about their publications before the indicator values are calculated. If the analysis is done on anything below university level it is particularly important.

CWTS indicators and denotations are included in this indicator definition list where appropriate, since these are well known in the bibliometric community. We have also chosen to include some indicators developed by the Swiss CEST, together with their denotations.

Note: The word *unit* is here to be interpreted as “unit of analysis”, unless in the context of “research unit”.

Indicators

Denotation index

P	Total number of publications
P_{ISI}	Number of publications in Thomson ISI indices
P_{TJ}	Number of publications in top journals
$P_{f5\%}$	Number of articles among the top 5% most cited in the field, of the same age and article type
p	Relative share of publications
$p_{f5\%}$	Top 5% – share of articles among top 5% most cited in the field, of the same type and age
p_u	Uncitedness – share of uncited publications
p_x	Co-authoring – share of publications co-authored with another unit
p_w	CEST field-based world share of publications
C	Total number of citations
c_i	Number of citations to a single publication i
$\bar{c}$	Average number of citations per publication
$\bar{c}_f$	Item oriented field normalized citation score average
C_f	Total item oriented field normalized citation score
$\left[\bar{c}\right]_f$	CWTS field normalized citation score (crown indicator)
$\bar{c}_{fz[\ln]}$	Item oriented field normalized logarithm-based citation z-score average
$\left[\bar{c}\right]_j$	Journal normalized citation score
$\bar{c}_j$	Item oriented journal normalized citation score average
$\left[\bar{c}\right]_{jp}$	Journal packet citation score
c_s	Self citedness – share of citations from the own unit
μ_f	Field reference value (field citation score) for articles of the same type, age and in the same field of research
$\bar{\mu}_f$	Mean field reference value (mean field citation score)
$\tau_{f5\%}$	Top 5% threshold value for the field; i.e. articles of the same type and age in the same scientific field
$\mu_{f5\%}$	Top 5% reference value for the field; i.e. articles of the same type and age in the same scientific field
$\mu_{f50\%}$	Top 50% reference value for the field; equals the median of the field
μ_j	Journal reference value
h	h-index
I_{ISI}	ISI journal impact factor
I_f	Journal to field impact score
ι_f	Field reference value for journals, based on a specified time window

Abbreviations

CWTS Center for Science and Technology Studies, Leiden University

ISI Thomson Scientific, formerly known as Thomson ISI

CEST Centre d'études de la science et de la technologie, Switzerland

Publications

Number of publications

Designation	**Total number of publications**
Denotation	P
Description	The number of scientific publications produced by the analyzed unit during the analyzed time span. Sometimes results are also presented separately per document type.
Calculation	Count the full number of scientific publications produced at the analyzed unit during the analyzed time span.
Formula	-
Example	-
Data Requirements	Verified publication data from a local publication source or the Thomson ISI indices complemented by self-reported publications from the analyzed unit.
Advantages	Relatively easy to produce.
Disadvantages	Does not take the size of the analyzed unit into account and does not say anything about the impact of the publications.
KI usage	At Karolinska Institutet we currently give every contributing unit full credit for the publication, i.e. no fractionalization or weighting between authors or institutions is used.
Reference	-

Number of ISI publications

Designation	**Number of publications in Thomson ISI indices**
Denotation	P_{ISI}
Description	The number of scientific publications produced by the analyzed unit during the analyzed time span, found in the Thomson ISI indices.
Calculation	Count the full number of publications in Thomson ISI indices, produced at the analyzed unit during the analyzed time span.
Formula	-
Example	-
Data Requirements	Verified publication data from Thomson ISI indices.
Advantages	Easy to retrieve from the Thomson ISI Web of Science.
Disadvantages	Does not take the size of the analyzed unit into account and does not say anything about the impact of the publications. Does not take into account publications not present in the Thomson ISI indices.
KI usage	At Karolinska Institutet, publications where the analyzed unit is only a part

	contributor is fully accounted to the unit, i.e. no fractionalization or weighting between authors is currently used.
Reference	-

Number of publications in top journals

Designation	**Number of publications in top-ranked journals**
Denotation	P_{TJ}
Description	The number of publications the analyzed unit has published in a selected number of journals during the analyzed time span.
Calculation	Select journals according to a suitable criterion. Check how many of the unit's publications that are published in these journals during the analyzed time span.
Formula	-
Example	-
Data Requirements	A bibliographic database (for instance Thomson ISI Web of Science or local publication database) to count publications and addition of publications not present in the database.
Advantages	Does reflect the potential impact of the unit's articles more than a mere publication count.
Disadvantages	Does not take the size of the analyzed unit into account.
KI usage	At Karolinska Institutet, journals classified as being focused on other subjects than life sciences are sometimes excluded from the journal list to make the indicator more relevant for assessments of life science research. No fractionalization or weighting between authors is currently used. At Karolinska Institutet, we also sometimes limit the selection of journals to journals containing original research articles, that is not pure Review Journals.
Reference	-

CEST world share of publications

Designation	**CEST field-based world share of publications**
Denotation	p_w
CEST Denotation	Part mondiale des publications.
Description	The unit's number of publications in each subdomain (research field) where the unit is active (full address counting, fractional field counting) is divided by the total number of world publications in the corresponding subdomains.
Calculation	To get one comprehensive value for a whole unit, a mean value for the unit's share of world publications is calculated, according to CEST (pg. A13 in ref. mentioned below): The unit's share of publications in each subdomain (field) is multiplied by the worldwide number of publications in each of the corresponding subdomains and the sum of the values retrieved from all subdomains is then divided by the sum of world publications in all subdomains where the unit is active. The result is multiplied with thousand and thus presented as a per mille number.
Formula	$p_w = 1000 \frac{\sum_{i=1}^{F} [p_f]_i [P_f]_i}{\sum_{i=1}^{F} [P_f]_i}$ KI suggested alternative formula: $p_w = 1000 \frac{\sum_{i=1}^{F} [P_{Uf}]_i}{\sum_{i=1}^{F} [P_f]_i}$ where: $[p_f]_i$ = the unit's share of publications in field i $[P_{Uf}]_i$ = the unit's number of publications in field i $[P_f]_i$ = the world total number of publications in field i F = the number of fields where the analyzed unit is active p_w = the CEST field-based world share of publications
Example	-
Data Requirements	Requires data from a comprehensive bibliographic database such as the Thomson citation indices.
Advantages	
Disadvantages	This indicator is not normalized with regard to document type or publication year. The classification used for domains and subdomains is the journal classification scheme supplied by Thomson Scientific.
KI usage	At KI this indicator is not used at present.
Reference	Annexe: aspects méthodologiques. (2004). Retrieved October 7th, 2006, from http://www.cest.ch/Publikationen/2004/method_2004.pdf

CEST degree of specialization

Designation	**CEST degree of specialization**

Denotation	-
CEST Denotation	Degré de specialisation
Description	The degree of specialization is a structural indicator that is affected by the number of subdomains in which a unit is active and how many publications there are in each of these. The degree of specialization for the whole world is by definition 0. A very specialized unit can have a maximum degree of specialization of 1. Between these two extremes there are 5 classes: <0.2 Very low degree of specialization >=0.2 & <0.4 Low degree of specialization >=0.4 & <0.6 Medium degree of specialization >=0.6 & <0.8 High degree of specialization >=0.8 Very high degree of specialization
Calculation	In each of the 107 subdomains (ISI journal fields), the number of publications from a unit is divided with the total number of publications from that unit, even if the number of publications from the unit in some subdomains is zero. The same procedure is then done for the rest of the world, and the unit's ratios are divided respectively with the world ratios. The 107 results are normalized on a scale from -100 to 100 where 0 corresponds to the world average (by multiplying the results with 100 and subtracting 100 from the derived results). The results are then added together and the sum divided by $107*(100)^2$ to get the degree of specialization.
Formula	-
Example	-
Data Requirements	Requires data from a comprehensive bibliographic database such as the Thomson citation indices.
Advantages	
Disadvantages	This indicator is not normalized with regard to document type or publication year. The classification used for domains and subdomains is the journal classification scheme supplied by Thomson Scientific.
KI usage	At KI this indicator is not used at present.
Reference	Annexe: aspects méthodologiques. (2004). Retrieved October 7th, 2006, from http://www.cest.ch/Publikationen/2004/method_2004.pdf

CEST relative activity index

Designation	**CEST relative activity index**
Denotation	-

CEST Denotation	"Activité" or RAI (Indice relatif de publication)
Description	The CEST relative activity index describes if a unit is more or less active in their chosen subdomains than the rest of the world.
Calculation	The publication count is fractionalized with regard to subject, i.e. an article categorized in three journal classes contributes 1/3 of an article to each subject. The number of a unit's publications in a particular subdomain (full address counting) is divided by the total number of publications from that unit. The same procedure is then done for the rest of the world. The share of the unit's publications is then divided by the share of the world's publications. To produce the RAI you then normalize the value to a scale of 0-200 where 100 equals the world average.
Formula	Publications in subdomain for unit Y/Total publications from unit Y = PY Publications in subdomain for the world/Total publications in the World = PW PY/PW = p RAI = $100 + 100 * (p^2-1)/(p^2+1)$
Example	-
Data Requirements	Requires data from a comprehensive bibliographic database such as the Thomson citation indices.
Advantages	-
Disadvantages	The indicator is not normalized with regard to document type or publication year. The classification used for domains and subdomains is the journal classification scheme supplied by Thomson Scientific.
KI usage	At KI this indicator is not used at present.
Reference	Annexe: aspects méthodologiques. (2004). Retrieved October 7th, 2006, from http://www.cest.ch/Publikationen/2004/method_2004.pdf

Citations

Number of citations

Designation	**Total number of citations**
Denotation	C
Description	The total number of citations to articles published by an analyzed unit during the analyzed time span.
Calculation	Find all articles published by the analyzed unit during the analyzed time span and sum their citation values (usually retrieved from the Thomson ISI indices via Web of Science or another source to the indices).
Formula	$C = \sum_{i=1}^{P} c_i$ where: P = number of publications c_i = number of citations for publication i
Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
Advantages	Gives an indication of the scientific impact the unit's published articles as a whole.
Disadvantages	Does not take into account that older articles usually are more cited and that citation rates vary between document types and subject areas. Does not compensate for the size of the unit.
KI usage	At Karolinska Institutet, publications where the analyzed unit is only a part contributor of the publication is fully accounted to the unit, and thus also the corresponding citation count.
Reference	-

Citations per publication

Designation	**Average number of citations per publication**
Denotation	$\bar{c}$
CWTS Denotation	CPP
Description	The average number of citations to articles published by an analyzed unit during the analyzed time span.
Calculation	Find all articles published by the analyzed unit during the analyzed time span in a citation index, sum up the citations and divide by the number of publications.

Formula	$\bar{c} = \frac{1}{P}\sum_{i=1}^{P} c_i$ where: c_i = number of citations for publication i P = number of publications
Example	-
Data Requirements	A comprehensive citation index as Thomson ISI citation indices and verification of the unit's articles.
Advantages	Gives an indication of the average scientific impact of the unit's published articles.
Disadvantages	Does not take into account that older articles usually are more cited if a variable, cumulative citation time window is used, and that citation rates vary between document types and subject areas.
KI usage	At Karolinska Institutet, citations to publications where the analyzed unit is only a part contributor of the publication is fully accounted to the unit.
Reference	-

CWTS field normalized citation score (crown indicator)

Designation	**CWTS field normalized citation score**
Denotation	$\left[\bar{c}\right]_f$
CWTS Denotation	CPP/FCSm or "crown indicator"
Description	This indicator corresponds to the number of citations to publications from a specific unit during an analyzed time span, compared to the world average of citations to publications of the same document types, ages and subject areas, seen as a group. The normalization of citation values is done on the sums of the citations and the field citation scores. The indicator is stated as a decimal number that shows the relation of the indicator to the world average, 1. As an example, 0.9 means that the unit's publications are cited 10% below average and 1.2 that they are cited 20% above average.
Calculation	Count all citations to the unit's publications and add them together. Add together all the world citation averages that correspond to the selected publications with respect to document type, publication year and research area. Divide the sum of citations with the sum of world averages.
Formula	$\left[\bar{c}\right]_f = \frac{\sum_{i=1}^{P} c_i}{\sum_{i=1}^{P}\left[\bar{\mu}_f\right]_i}$

	where: c_i = number of citations to publication i $\left\lfloor \bar{\mu}_f \right\rfloor_i$ = the average value of citations to publications of the same type, published the same year in the same research areas as article i P = number of publications
Example	A research unit has published three articles: • One original article A was published in 2000 within research area X. This has received 9 citations. • One review B was published in 2001 within research area Y. This has received 21 citations. • One original article C was published in 2002 within research area Z. This has received 4 citations. The field citation scores for corresponding articles are: • Original articles published in 2000 in research area X= 5.2 citations • Review articles published in 2001 in research area Y = 26.3 citations • Original articles published in 2002 in research area Z= 3.2 citations The citation values and the field citation scores are added together before normalization: (9+21+4) / (5.2+26.3+3.2) = 0.98. A CWTS field normalized citation score of 0.98 means that the unit's publications are cited 2 % below average. *[Note that for reasons of clarity the number of publications in this example is much lower than the minimum value recommended for a bibliometric analysis.]*
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
Advantages	By summing the citations from all fields before normalization, this indicator is relatively resistant to citation scores being levered by a few highly cited papers in low-cited fields.
Disadvantages	Citation rates are not normalized on the level of individual publications, but on a higher aggregation level where the average citation rate of a researcher, group or department is compared to the average citation rate of the fields in which the researcher or group has published. This way of calculating gives more weight to older publications (particularly reviews), published in fields with dense citation traffic.
KI usage	At KI, the indicator "item oriented field normalized citation score average" is used as an alternative to the CWTS crown indicator.
Reference	Moed, H. F., Debruin, R. E., & Vanleeuwen, T. N. (1995). New Bibliometric Tools for the Assessment of National Research Performance - Database Description, Overview of Indicators and First Applications. Scientometrics, 33(3), 381-422.

Field normalized citation score

Designation	**Item oriented field normalized citation score average**
Denotation	$\bar{c}_f$
Description	This indicator corresponds to the relative number of citations to publications from a specific unit, compared to the world average of citations to publications of the same document type, age and subject area. The term "item oriented" indicates that the normalization of the citation values is done on an individual article level. This indicator is stated as a decimal number that shows the relation of the number of citations to the world average. As an example, 0.9 means that a unit's publications are cited 10% below average and 1.2 that they are cited 20% above average.
Calculation	The number of citations to each of the unit's publications is normalized by dividing it with the world average of citations to publications of the same document type, publication year and subject area, which is called the field reference value (μ_f). If an article is classified as belonging to several subject areas, a mean value of the areas is used. The indicator is the mean value of all the normalized citation scores for the unit's publications.
Formula	$\bar{c}_f = \frac{1}{P}\sum_{i=1}^{P}\frac{c_i}{\left[\bar{\mu}_f\right]_i}$ where: c_i = number of citations to publication i $\left[\bar{\mu}_f\right]_i$ = the average value of citations to publications of the same type, published the same year in the same research area as article i P = the unit's number of publications
Example	A research unit has published three articles: • One original article A was published in 2000 within research area X. This has received 9 citations. • One review B was published in 2001 within research area Y. This has received 21 citations. • One original article C was published in 2002 within research area Z. This has received 4 citations. The field reference values (μ) for corresponding articles are: • Original articles published in 2000 in research area X= 5.2 citations • Review articles published in 2001 in research area Y = 26.3 citations • Original articles published in 2002 in research area Z= 3.2 citations The field normalized citation score for each article is: • A (Original Article/X/2000): 9/5.2 = 1.73 • B (Review/Y/2001): 21/26.3 = 0.80

	• C (Original Article Z/2002): 4/3.2 = 1.25 The average of the normalized citation scores is: (1.73 + 0.80 + 1.25) / 3 = 1.26. The item oriented field normalized citation score for this unit is 1.26 which means that publications from this research unit are cited 26 % above average. Note that the original article from 2000 is the main contributor to this high value although the review has received more citations. *[Note that for reasons of clarity the number of publications in this example is much lower than the minimum value recommended for a bibliometric analysis.]*
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices and calculation of field normalized citation scores for normalization of citation values.
Advantages	As the normalization takes place on the level of the individual publication the indicator gives each publication equal weight in the final value.
Disadvantages	If the normalization is done on an article level, a few highly cited articles in a moderately cited research area may contribute unproportionately to the value of the field normalized citation score.
KI usage	At KI, the item oriented field normalized citation score average is used as an alternative to the CWTS Crown Indicator. Citations to publications where the analyzed unit is only a part contributor of the publication is fully accounted to the unit.
Reference	Lundberg, J: Lifting the crown – Citation Z-score. Journal of Informetrics (submitted).

Total field normalized citation score

Designation	**Total item oriented field normalized citation score**
Denotation	C_f
Description	This indicator gives an indication of both the impact and the production volume of the analyzed unit.
Calculation	Add together the item oriented field normalized citation scores for all the publications of the analyzed unit.
Formula	$C_f = \sum_{i=1}^{P} \left\lfloor c_f \right\rfloor_i$ where: $\left\lfloor c_f \right\rfloor_i$ = item oriented field normalized citation score for publication i P = total number of publications for the analyzed unit
Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices and calculation of field normalized citation scores for normalization of citation values.
Advantages	Gives and indication on both the volume and the impact of the publications from

	the analyzed unit.
Disadvantages	Does not compensate for the size of the analyzed unit.
KI usage	
Reference	Lundberg, J: Lifting the crown – Citation Z-score. Journal of Informetrics (submitted).

Logarithm-based citation z-score

Designation	**Item oriented field normalized logarithm-based citation z-score average**
Denotation	$\bar{c}_{fz[\ln]}$
Description	The logarithm-based citation z-score relates the logarithm of the number of citations that a publication has received with to the mean and the standard deviation for the logarithms of the citation rates for all the corresponding reference publications of the same type, age and subject area.
Calculation	The average of the logarithms of the number of citations (plus 1 to avoid the value 0) to publications of the same document type, publication year and subject area, which is called the logarithm-based field reference value ($\mu_{f[\ln]}$), is subtracted from the logarithms of the citation counts (plus 1) for each article produced by the analyzed unit during the analyzed time span. If an article is classified as belonging to several subject areas, a mean value of the areas is used as $\mu_{f[\ln]}$. The resulting value is then divided by the standard deviation for the logarithm of the citation count plus one of the population of articles that constitutes the logarithm-based field reference value. Finally, the mean value of all values calculated as mentioned above is calculated by dividing the values with the number of analyzed publications, and this gives the logarithm-based citation z-score indicator for the unit.
Formula	$\bar{c}_{fz[\ln]} = \frac{1}{P}\sum_{i=1}^{P}\frac{\ln(c_i+1)-\lfloor\mu_{f[\ln]}\rfloor_i}{\lceil\sigma_{f[\ln]}\rceil_i}$ where: c_i = number of citations to publication i $[\mu_{f[\ln]}]_i$ = the logarithm-based field reference value; the average value of the logarithms of the number of citations plus one to publications of the same type, published the same year in the same research area as article i $[\sigma_{f[\ln]}]_i$ = the standard deviation of the $[\mu_{f[\ln]}]_i$ distribution P = the unit's number of publications
Example	If a review article published in 2000 in Nature Reviews Immunology has received 66 citations. The logarithm of this value plus one (4.2) would then be compared with the average number (2.7) and standard deviation (1.3) of the logarithms of citation rates (plus one) of all reviews from 2003 in immunology. The citation z-score for this article is then (4.18-2.7) / 1.3 = 1.1. Observe that the comparison is made with average of the logarithms of the number of citations received by comparable items and not with the logarithm of the

	average number of citations received by comparable items. The bibliometric indicator for a research group, department or university is then the item oriented field normalized logarithm-based citation z-score average. The citation z-score could for instance be something like (2.1+1.0+1.1+0.5+1.0)/5=1.1. The publications in the example are thus, after logarithmic transformation, on average cited 1.1 standard deviations above the world average for publications of the same type, from the same year, published in journals belonging to the same subject category.
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices and verification of articles as belonging to the analyzed unit.
Advantages	Since the distribution of citation rates differs between research fields, publication years and document types it can be argued that using a z-score in the normalization procedure would be more appropriate than a field normalized citation score. The z-score indicator gives information both if the citation value of the publication is lower (negative z-score) or higher (positive z-score) than the field score, and how far from the mean the value is, measured in a normalized way by using the standard deviation for the field citation score as a measuring unit.
Disadvantages	Rather complicated to calculate. If the distribution of citation values in the publications of the field is much skewed and thus far of from a normal distribution, both the mean value and the standard deviation may be somewhat misleading measures to use for an indicator.
KI usage	This indicator is presently being developed and refined by bibliometric researcher Jonas Lundberg at Karolinska Institutet.
Reference	Lundberg, J: Lifting the crown – Citation Z-score. Journal of Informetrics (submitted).

Top 5%

Designation	**Top 5%**
Denotation	$p_{f5\%}$
CWTS denotation	A/E (Ptop)
Description	Top 5% shows the share of publications attributed to a unit that belong to the 5% most cited publications in the world from the same year, in the same subject and of the same document type. Other top values, as top 1% and top 10% are also used, and calculated in the same way as top 5%. The indicator is written as a decimal number that shows the relation to the world average. A value over 1 shows that the analyzed unit has more of its publications among the top 5% than the world average, a value below 1 that it has less.
Calculation	Find the number of citations needed for a publication to belong to the 5% most cited publications of the same document type, with the same publication year and in the same research area ($\tau_{f5\%,}$ see Top 5 % reference values). If the article is classified as belonging to several subject areas, a mean value between subject areas is calculated.

	Find the share of publications in the world above that threshold value within the same document type, with the same publication year and in the same research area ($\mu_{f5\%}$ see Top 5 % reference value). Since we apply a strict rule that the number of citations need to be above the threshold value and since the distribution of citation is skewed, the world share within a group of publications is not always = 0,05. Count how many of the analyzed publications that have more citations than the threshold value, $\tau_{f5\%}$, found above. Each publication must be compared with the threshold value for publications of the same document type, with the same publication year and within the same research area, or a mean value if the article is classified as belonging to several subject areas. Divide the number of analyzed publications with a citation value above the threshold with the total number of analyzed publications. Divide the received value with $\mu_{f5\%}$ (the mean value of the world average) to get the value of the Top 5 % indicator.
Formula	$p_{f5\%} = (P_{f5\%} / P) / \mu_{f5\%}$ where: $P_{f5\%}$ = number of publications above citation threshold for 5% most cited for the same article type, year and field P = total number of publications for the analyzed unit during the analyzed time span $\mu_{f5\%}$ = the mean world share of publictions above citation threshold for 5% most cited for the same article type, year and field
Example	A research unit has published three articles: • One original article A was published in 2000 within research area X. This has received 9 citations. • One review B was published in 2001 within research area Y. This has received 103 citations. • One original article C was published in 2002 within research area Z. This has received 4 citations. To belong to the 5% most highly cited: • Original articles published in 2000 in research area X: 8 citations • Review articles published in 2001 in research area Y: 103 citations • Original articles published in 2002 in research area Z: 36 citations Are the publications of this research group among the 5% most highly cited? X: 9 > 8; Yes Y: 103 > 103; No Z: 4 > 36; No Share of publications among the 5% most highly cited ($P_{f5\%}$/P): 1/3 = 33% Mean value of the share of publications in the world among the 5% most highly cited ($\mu_{f5\%}$):

	$$\mu_{f5\%} = \frac{\sum_{i=1}^{P} N_i \left\lfloor \mu_{f5\%} \right\rfloor_i}{\sum_{i=1}^{P} N_i}$$ Where: $[\mu_{f5\%}]_i$ = the share of publications in the world among the 5% most highly cited with the same article type, year and field as the publication i N_i= number of publications in the world in each group of publications with the same article type, year and field as the publication i The Top 5% indicator has a value of 0.33/0.04= 8.25 Often the indicator ($P_{f5\%}$ / P) is used instead which in this example equals 0.33 i.e. the share of publications in the unit with more citations than the 95_{th} percentile limit. *[Note that for reasons of clarity the number of publications in this example is much lower than the minimum value recommended for a bibliometric analysis.]*
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
Advantages	Gives an indication about the share of very top impact publication and can be used to augment the field normalized citation score to reveal if a high normalized citation score is due to few highly-cited articles or a general high level of citations to the unit's articles.
Disadvantages	
KI usage	
Reference	van Leeuwen T.N., V. M. S., Moed H.F. and Nederhof A.J. . (2002). Bibliometric Profiles of Academic Chemistry Research in the Netherlands,1991 - 2000.

CEST relative impact index

Designation	**CEST relative impact index**
Denotation	-
CEST Denotation	"Impact" or RZI (indice relatif de citation)
Description	Gives an indication of the relative "audience" for publications from the analyzed unit compared to the world average. It is counted separately for each subdomain.
Calculation	Full counting is done on fields but fractional counting on addresses and citations. The citation count is fractionalized with regard to the length of the reference list, for example, if a reference list contains 14 references each cited article will receive 1/14 of a citation. The average number of citations per publication in each subdomain is counted for the articles from the unit of analysis. These values are then divided with the average number of citations per publication for international publications in the

	corresponding subdomains. To produce the RZI the value is then normalized to a scale of 0-200 where 100 equals the world average.
Formula	Citations to P in subdomain unit Y/ Publications in subdomain unit Y = CPPY Citations to P in subdomain World / Publications in subdomain world = CPPW CPPY/PPW = i RZI=100 + 100 * $(i^2-1)/(i^2+1)$
Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
Advantages	
Disadvantages	This indicator is not normalized with regard to document type or publication year. The classification used for domains and subdomains is the journal classification scheme supplied by Thomson Scientific.
KI usage	At KI this indicator is not used at present.
Reference	Annexe: aspects méthodologiques. (2004). Retrieved October 7th, 2006, from http://www.cest.ch/Publikationen/2004/method_2004.pdf

CEST normalized mean impact

Designation	**CEST normalized mean impact**
Denotation	-
CEST Denotation	L'impact moyens pondéré
Description	This indicator is based on number of publications and the CEST relative impact indicators for each subdomain in which a unit is active. 0-40 : Very low 40-80 Low 80-120 Medium 120-160 High 160-200 Very high
Calculation	The unit's number of publications in each subdomain where the unit is active is multiplied by the CEST relative impact indicator for the corresponding subdomain. All the resulting values are then added together and divided by the total number of publications from the analysed unit in the subdomains where they are active.
Formula	-
Example	
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.

Advantages	
Disadvantages	This indicator is not normalized with regard to document type or publication year. The classification used for domains and subdomains is the journal classification scheme supplied by Thomson Scientific.
KI usage	At KI this indicator is not used at present.
Reference	Annexe: aspects méthodologiques. (2004). Retrieved October 7th, 2006, from http://www.cest.ch/Publikationen/2004/method_2004.pdf

CWTS journal normalized citation score

Designation	**CWTS Journal normalized citation score**
Denotation	$\left[\bar{c}\right]_j$
CWTS Denotation	CPP/JCSm
Description	This indicator corresponds to the number of citations to publications from a specific unit during an analyzed time span, compared to the world average of citations to publications of the same document types, ages and in the same journals, seen as a group. The normalization of citation values is done on the sums of the citations and the journal citation scores. The indicator is stated as a decimal number that shows the relation of the indicator to the world average, 1. As an example, 0.9 means that the unit's publications are cited 10% below average and 1.2 that they are cited 20% above average. A high indicator value suggests that a group is highly cited within the journals they chose to publish in.
Calculation	Count all citations to the unit's publications and add them together. Add together all the world averages that correspond to the selected publications with respect to document type, publication year and in the same journals. Divide the sum of citations with the sum of world averages.
Formula	$\left[\bar{c}\right]_j = \frac{\sum_{i=1}^{P} c_i}{\sum_{i=1}^{P} \left[\mu_j\right]_i}$ where: c_i = number of citations to publication i $[\mu_j]_i$ = the average value of citations to publications of the same type, published the same year in the same journal as article i P = number of publications
Example	A research unit has published three articles: • One original article A was published in 2000 in journal X. This has received 9 citations.

	• One review B was published in 2001 in journal Y. This has received 21 citations. • One original article C was published in 2002 in journal Z. This has received 4 citations. The field citation scores for corresponding articles are: • Original articles published in 2000 in journal X= 5.2 citations • Review articles published in 2001 in journal Y = 26.3 citations • Original articles published in 2002 in journal Z= 3.2 citations The citation values and the journal citation scores are added together before normalization: (9+21+4) / (5.2+26.3+3.2) = 0.98. A CWTS journal normalized citation score of 0.98 means that the unit's publications are cited 2 % below average. *[Note that for reasons of clarity the number of publications in this example is much lower than the minimum value recommended for a bibliometric analysis.]*
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices and verification of publications produced by the unit.
Advantages	Since the normalization of the citation score is made on the sums of the citations and journal citation scores, the CWTS Journal Normalized Citation Score is relatively resistant to if only a few of the publications from a unit of analysis have a very high or low citation count compared to the corresponding world average.
Disadvantages	Citation rates are not normalized on the level of individual publications, but on a higher aggregation level where the average citation rate of a researcher, group or department is compared to the average citation rate of the fields in which the researcher or group has published. It is possible to manipulate this indicator by a adopting a strategy aimed at publishing averagely cited articles in journals with a below average journal impact indicator. This is however easy to discover if you combine this indicator with the journal packet citation score indicator.
KI usage	At Karolinska Institutet, this indicator is not used at present.
Reference	Moed, H. F., Debruin, R. E., & Vanleeuwen, T. N. (1995). New Bibliometric Tools for the Assessment of National Research Performance - Database Description, Overview of Indicators and First Applications. Scientometrics, 33(3), 381-422.

Journal normalized citation score

Designation	**Item oriented journal normalized citation score average**
Denotation	$\bar{c}_j$
CWTS Denotation	-
Description	This indicator corresponds to the number of citations to publications from a specific unit during an analyzed time span, compared to the world average of citations to publications of the same document types,

	ages and in the same journals. The term "item oriented" indicates that the normalization of the citation values is done on an individual article level. The indicator is stated as a decimal number that shows the relation of the indicator to the world average, 1. As an example, 0.9 means that the unit's publications are cited 10% below average and 1.2 that they are cited 20% above average. A high indicator value suggests that a group is highly cited within the journals they chose to publish in.
Calculation	The number of citations to each of the unit's publications is normalized by dividing it with the world average of citations to publications of the same document type, published the same year in the same journal. The indicator is the mean value of all the normalized citation counts for the unit's publications.
Formula	$\overline{c_j} = \frac{1}{P}\sum_{i=1}^{P}\frac{c_i}{\mu_j}$ c_i = number of citations to publication i μ_j = the average number of citations to publications of the same type, published the same year and in the same journal area as article i P = the unit's number of publications
Example	A research unit has published three articles: • One original article A was published in 2000 in journal X. This has received 9 citations. • One review B was published in 2001 in journal Y. This has received 21 citations. • One original article C was published in 2002 in journal Z. This has received 4 citations. The journal reference values for corresponding articles are: • Original articles published in 2000 in journal X= 5.2 citations • Review articles published in 2001 in journal Y = 26.3 citations • Original articles published in 2002 in journal Z= 3.2 citations The journal normalized citation scores for each article are: • A (Original Article/X/2000): 9/5.2 = 1.73 • B (Review/Y/2001): 21/26.3 = 0.80 • C (Original Article /Z/2002): 4/3.2 = 1.25 The average of the normalized citation scores is: (1.73 + 0.80 + 1.25) / 3 = 1.26. An item oriented journal normalized citation score of 0.98 means that the unit's publications are cited 26 % above average. *[Note that for reasons of clarity the number of publications in this example is much lower than the minimum value recommended for a*

	bibliometric analysis.]
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
Advantages	
Disadvantages	
KI usage	At Karolinska Institutet, this indicator is not used at present.
Reference	-

Journal packet citation score

Designation	**Journal packet citation score**
Denotation	$\left[\bar{c}\right]_{jp}$
CWTS denotation	JCSm/FCSm
Description	The average impact of the journals in which a unit has published relative to the world average in the fields covered by this set of journals. If the value is above one the unit has published in journals with relatively high impact.
Calculation	-
Formula	$\left[\bar{c}\right]_{jp} = \frac{\sum_{i=1}^{P} c_i}{\sum_{i=1}^{P} \left[\mu_j\right]_i}$ where: c_i = number of citations to publication i $[\mu_j]_i$ = the average value of citations to publications of the same type, published the same year in the same journal packet as article i P = number of publications
Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
Advantages	-
Disadvantages	-
KI usage	At Karolinska Institutet, this indicator is not used at present.
Reference	Van Leeuwen, T. N., & Moed, H. F. (2002). Development and application of journal impact measures in the Dutch science system. Scientometrics, 53(2), 249-266.

h-index

Designation	**Hirsch index (h-index)**
Denotation	h
CWTS denotation	h
Description	The h-index is the number of publications (h), attributed to the analyzed unit during the analyzed time span, that have at least h citations.
Calculation	Find the unit's published articles in a citation index and sort them in descending order by number of citations. Count articles from the top of the list and downwards, and when the number of an article rises above the citation count for that very article, the number of the preceding article is to be counted as the h-index.
Formula	See Hirsch's original article, referenced below.
Example	According to the Web of Science (WoS), a unit has published 169 articles during the analyzed time span. The articles are sorted in descending citation count order in WoS and it is found that article number 32 has 33 citations and article number 33 has 31 citations, which is lower than the article number. The h-index will therefore be 32, since the unit thus has 32 articles with at least 32 citations.
Data Requirements	A comprehensive citation index as Thomson ISI citation indices.
Advantages	Very easy to calculate in the ISI Web of Science.
Disadvantages	h-index gives positive bias to senior researchers with older articles, since these have had more time to be cited, though the demand that new articles with comparable citation levels has to be added has a certain damping effect on that bias.
KI usage	h-index is presently not used by the Karolinska Institutet bibliometrics group.
Reference	Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569-16572

Uncitedness

Designation	**Uncitedness**
Denotation	p_u
CWTS Denotation	%Pnc
Description	The share of a unit's publications that that remain uncited after a certain time period. Self-citations should be removed from the citation count.
Calculation	Count the number of publications that have never been cited during a specified time period, excluding self-citations. Divide with the total number of publications from the same unit during the same time period.

Formula	$p_u = P_u / P$ where: P_u = the unit's number of publications which has received no citations P = the unit's total number of publications
Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices and validation of the unit's publications.
Advantages	-
Disadvantages	-
KI usage	
Reference	-

Self citedness

Indicator	**Self citedness**
Denotation	c_s
CWTS denotation	%SELFCIT
Description	The share of a unit's received citations where authors refer to their own papers.
Calculation	Count the total number of citations to the unit's publications during the analyzed time span. Check where citations are coming from and count the number coming from the unit itself. Divide the second number with the first to get share of self citedness.
Formula	$c_s = C_S / C$ where: C_S = citation to the unit's publications emanating from the unit itself C = the total number of citations to the unit's publications
Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices, validation of publications and analysis of citing articles, which can be done in the ISI Web of Science.
Advantages	-
Disadvantages	-
KI usage	
Reference	-

Cooperation

Co-authoring

Denomination	**Share of articles co-authored with another unit**
Denotation	p_x
Description	This group of indicators is used to show to what extent an analyzed unit cooperates with other units in the production of articles. • International collaboration – share of publications with co-authors from organizations in at least two different countries. • National collaboration – share of publications with co-authors from at least two organizations within the same country. • Department collaboration – share of publications with co-authors from at least two departments within the same organization. • Unit collaboration - share of publications with co-authors from two or more research units.
Calculation	Count the number of articles published by the analyzed unit during the analyzed time span and check how many that was co-authored together with a selected other unit. Divide the second figure by the first one to get the share of articles co-authored between the units.
Formula	$p_x = P_x / P$ where: p_x = share of publications co-authored with a certain unit P_x = number of publications co-authored with the selected unit P = total number of publications produced at the analyzed unit during the analyzed time
Example	-
Usage	
Data Requirements	Verified article data and full addresses to all participating units.
Advantages	
Disadvantages	
KI Usage	
Reference	
KI Usage	
Reference	

Journals

ISI journal impact factor

Designation	**ISI Journal Impact Factor**
Denotation	I_{ISI}
CWTS denotation	IF
Usage	Used to measure the impact of scientific journals.
Description	The ISI impact factor is a number that corresponds to the average number of citations a publication in a specific journal has received during the two years following the year of publication.
Calculation	The ISI impact factor for a specific journal (J), one specific year (Y) is calculated by counting the number of citations to articles in that journal the two preceding years (Y-1 and Y-2) from publications in year Y and dividing this with the number of publications defined by Thomson ISI as "citeable" in journal J the two preceding years (Y-1 and Y-2).
Formula	$I_{ISI} = C / P$ where: I_{ISI} = the impact factor for journal J in year Y C = the number of citations from publications in year Y to publications in journal J published Y-2 and Y-1 P = total number of citeable publications in journal J in year Y-2 and Y-1
Example	The 2005 impact factor of the journal Nature is produced by counting the number of citeable publications in Nature during 2005 that cite publications in nature from 2003-2004 and dividing this with the total number of publications in Nature 2003-2004.
Data Requirements	No own data is required; ISI journal impact factor is available through the ISI service Journal Citation Reports.
KI Usage	
Reference	THE ISI IMPACT FACTOR by Thomson Scientific: http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/

Normalized journal impact

Designation	**Normalized journal impact**
Denotation	$\bar{c}_f$
Description	Equal to an item oriented field normalized citation score for articles from only one journal. This indicator corresponds to the relative number of citations to publications in one specific journal, compared to the world average of citations to publications of the

	same document type, age and subject area. The indicator is stated as a decimal number that shows the relation of the number of citations to the world average. As an example, 0.9 means that publications in this journal are cited 10% below average and 1.2 that they are cited 20% above average.
Calculation	The number of citations to each of the journal's publications is normalized by dividing it with the world average of citations to publications of the same document type, publication year and subject area, which is called the field citation score (μ_f). If an article is classified as belonging to several subject areas, the mean value of the field citation scores is used. The indicator is the mean value of all the normalized citation counts for publications in this journal.
Formula	$\bar{c}_f = \frac{1}{P}\sum_{i=1}^{P}\frac{c_i}{\left[\bar{\mu}_f\right]_i}$ c_i = number of citations to publication i $\left[\bar{\mu}_f\right]_i$ = the average value of citations to publications of the same type, published the same year in the same research area as article i P = the number of publications in the journal during the selected time period
Example	In the year 2002 Journal J which belongs to research area Y and Z published three articles. The normalized journal impact is calculated in 2005 since most research areas reach their citation peak three years after publication. • Original article A which has received 9 citations. • Review B which has received 21 citations. • Original article C which has received 4 citations. The field citation scores (μ_f) for corresponding articles are: • Original articles published in 2002 in research area Y= 2.6 citations • Original articles published in 2002 in research area Z= 5.2 citations μ_f=(2.6+5.2)/2=7.8 • Review articles published in 2002 in research area Y = 11.7 citations • Review articles published in 2002 in research area Z = 26.3 citations μ_f=(11.7+26.3)/2=19.0 The field normalized citation score for each article is: • A: 9/7.8 =1.2 • B: 21/19 =1.1 • C: 4/7.8 =0.5 The average of the normalized citation scores is: (1.2+1.1+0.5) / 3 = 0.9 The 2002 normalized journal impact for Journal J calculated in 2005 is 0.9 which means that publications from 2002 published in this journal are cited 10% below average.

	[Note that for reasons of clarity the number of publications in this example is much lower than the minimum value recommended for a bibliometric analysis.]
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices and calculation of field normalized citation scores for normalization of citation values.
Advantages	
Disadvantages	
KI usage	
Reference	

Journal to field impact score

Designation	**Journal to field impact score**
Denotation	I_f
CWTS denotation	JFIS
Usage	Used to measure the relative impact of scientific journals.
Description	A more advanced journal impact factor than the ISI Impact Factor that takes both journal subject areas and document types into consideration. This makes comparison possible between journals in different subject areas. The Journal to Field Impact Score compares citations to one specific journal to the world average of citations to journals in the same subject area. An improvement of the ISI impact factor is to extend the period of measurement to for instance 5 years, since most articles have their citation peak 2-3 years after publication. A second improvement is to extend the ISI range of "citable publications" to include documents of type "letter", to make it more difficult to manipulate the impact score and to have the same publiciaton types in bothe the numerator and the denominator.
Calculation	-
Formula	$I_f = \frac{\bar{c}}{\bar{\imath}_f}$ where: I_f = the journal to field impact score for journal J in year Y $\bar{c}$ = the average number of citations from publications in year Y to publications in journal J published in year Y-1 to Y-5 $\bar{\imath}_f$ = the average number of citations to articles published in year Y in journals in the same fields as journal J in year Y-1 to Y-5

Example	In the year 2000 Journal J published 110 papers (counting only articles, letters and reviews). During 2000-2005 these publications were cited 1289 times. The average number of citations made in 2000-2005 to papers published 2000 in journal J (counting only articles, letters and reviews) is 1289/110 = 11.7 The world average of citations made 2000-2005 to papers (counting only articles, letters and reviews) published in 2000 in journals in the same field is 14.9. The Journal to Field Impact Score for journal J is 11.7/14.9=0.79 Journal J gets cited 21% below average.
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
KI Usage	Karolinska Institutet bibliometrics group strives at replacing the ISI Journal Impact Factor with the Journal to Field Impact Score whenever possible, since we believe it to be a better indicator on scientific journal performance, and thus more suitable for extrapolation of quality statements on recently published articles.
Reference	Van Leeuwen, T. N., & Moed, H. F. (2002). Development and application of journal impact measures in the Dutch science system. Scientometrics, 53(2), 249-266.

Citation reference values

Field citation reference value

Designation	**Field citation reference value**
Denotation	μ_f
CWTS Denotation	FCS (field citation score)
Description	The world average of citations to publications of the same document types, ages and subject areas.
Calculation	All documents are divided into groups where the items have the same document type, age and subject area. The mean value of the citations to all publications within the same group is the international field reference value for that particular group.
Formula	$\mu_f = \frac{1}{P}\sum_{i=1}^{P} c_i$ where: c_i = number of citations for publication i in field group P = number of publications in the field group
Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
KI usage	At Karolinska Institutet the field reference value is used to normalize citation rates for calculation of the more advanced citation indicators. Presently, we use the ISI subject classification of the journals where the articles were published as a basis for grouping articles by subject.
Reference	-

Top 5% citation reference values

Designation	**Field top 5% citation threshold value**
Denotation	$\tau_{f5\%}$
Description	The top 5% threshold value is the minimum number of citations essential to make a publication one of the 5% most cited publications of the same age, of the same publication type within the same field. Other top reference values, as top 1% and top 10% are also used, and calculated in the same way as top 5%.
Calculation	All publications are divided into groups where the items have the same document type, age and subject area. The publications in the group are counted and sorted according to the number of citations in descending order. The number of citations needed to belong to the top 5% share of publications, i.e. the 95_{th} percentile limit, is equal to the top 5% threshold value.
Formula	Se calculation above.

Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
KI usage	At Karolinska Institutet, the 5% reference value is used to calculate the Top 5% indicator.
Reference	-

Designation	**Field top 5% citation reference value**
Denotation	$\mu_{f5\%}$
Description	The top 5% reference value is the world share of publictions above citation threshold for 5% most cited for the same article type, year and field. Other top reference values, as top 1% and top 10% are also used, and calculated in the same way as top 5%.
Calculation	All publications are divided into groups where the items have the same document type, age and subject area. The publications in the group are counted as well as the number of publications in the group with citations above $\tau_{f5\%}$. The quota between those two numbers are calculated.
Formula	$\mu_{f5\% = P/Pf5\%}$
Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
KI usage	At Karolinska Institutet, the 5% reference value is used to calculate the Top 5% indicator.
Reference	-

Journal citation reference value

Designation	**Journal citation reference value**
Denotation	μ_j
CWTS Denotation	JCS (journal citation score)
Description	The world average of citations to publications in the same journal and of the same document type and age.
Calculation	All documents are divided into groups consisting of items published in the same journal, having the same document type and age. The mean value of the citations to all publications within the same group is the journal reference value for that particular group.
Formula	$\mu_j = \frac{1}{P}\sum_{i=1}^{P} c_i$

	where: c_i = number of citations to article i, belonging to the selected group of articles P = number of publications in the selected group of articles
Example	-
Data Requirements	Requires data from a comprehensive citation database such as the Thomson citation indices.
KI usage	
Reference	