ETH zürich

How to use Bibliometric Data to Rank Universities according to their Research Performance?

Rüdiger Mutz, ETH Zurich

COST Conference, Zurich, 12.2.-13.2.2015

- Hazelkorn (2013) differs 11 different international university rankings (e.g., Shanghai Ranking, Leiden Ranking)->see session "Ranking"
- Most important source of data: bibliometric data
- Why? Bibliometric data are easily available due to bibliographic databases (WOS, Scopus, …) and seem to be objective and reliable
- Social science methodology as a reference system for data analysis: measurement theory, hypothesis testing, statistical models, study design, theory-driven, operationalization of constructs, …

- Goldstein & Spiegelhalter (1996) named the key points for conducting quantitative comparison among institutions (i.e. league tables):

  "We shall pay particular attention to the specification of an appropriate statistical model, the crucial importance of uncertainty in the presentation of results, techniques for adjustment of outcomes for confounding factors and finally the extent to which any reliance may be placed on explicit rankings." (p. 390)

Four implications:

- A statistical instead of a numerical perspective on bibliometric data

- Bibliometric measures: $PP_{top10\%}$ instead of Crown indicators, full counting instead of fractional counting

- Confounding of bibliometric data: bias factors

- Visualizing of bibliometric data for ranking: Plots, maps

Example for a university ranking (fictional data)

University	$PP_{top10\%}$
A	0.267
C	0.174
B	0.150
D	0.139
E	0.131
G	0.126
F	0.112
Q	0.108
H	0.105
J	0.105
L	0.103
K	0.101
O	0.097
I	0.087
M	0.087
P	0.083
N	0.082
S	0.077
Z	0.058
R	0

Example for a university ranking (fictional data)

University	$PP_{top10\%}$ Population	Number of papers = N	Number of papers in top10% = y	$PP_{top10\%}$
A	0.1	15	4	0.267
C	0.1	86	15	0.174
B	0.1	80	12	0.150
D	0.1	122	17	0.139
E	0.1	168	22	0.131
G	0.1	190	24	0.126
F	0.1	241	27	0.112
Q	0.1	65	7	0.108
H	0.1	285	30	0.105
J	0.1	19	2	0.105
L	0.1	243	25	0.103
K	0.1	179	18	0.101
O	0.1	268	26	0.097
I	0.1	206	18	0.087
M	0.1	46	4	0.087
P	0.1	36	3	0.083
N	0.1	207	17	0.082
S	0.1	209	16	0.077
Z	0.1	137	8	0.058
R	0.1	13	0	0

y are random numbers, distributed according to a binomial distribution:

$y \sim \text{binomial}(N, p=0.1)$

Example of a confounding covariates: Sections of *Chemical Abstracts*

Correlation P and C of .57

Table 2. Bibliometric indicators for the biochemistry sections of *Chemical Abstracts*

Section		P	C	CPP
1.	Pharmacology	27,377	276,178	10.09
2.	Mammalian hormones	16,355	215,604	13.18
3.	Biochemical genetics	17,602	258,331	14.68
4.	Toxicology	10,184	68,569	6.73
5.	Agrochemical bioregulators	2,947	6,975	2.37
6.	General biochemistry	10,058	197,886	19.67
7.	Enzymes	9,573	129,583	13.54
8.	Radiation biochemistry	3,202	24,237	7.57
9.	Biochemical methods	8,320	76,694	9.22
10.	Microbial, algal, and fungal biochemistry	9,894	111,759	11.30
11.	Plant biochemistry	9,091	66,749	7.34
12.	Nonmammalian biochemistry	6,110	66,601	10.90
13.	Mammalian biochemistry	12,608	223,535	17.73
14.	Mammalian pathological biochemistry	22,941	312,522	13.62
15.	Immunochemistry	17,328	270,462	15.61
16.	Fermentation and bioindustrial biochemistry	2,675	10,510	3.93
17.	Food and feed chemistry	9,344	34,509	3.69
18.	Animal nutrition	4,994	34,637	6.94
19.	Fertilizers, soils, and plant nutrition	4,155	11,813	2.84
20.	History, education, and documentation	2,554	10,406	4.07

Note. P = number of research articles published in 2000, C = number of citations during the period 2000–2004 (including self-citations), CPP = average number of citations per research article.

Source: Neuhaus, C. & Daniel H.-D. (2009). A new reference standard for citation analysis in chemistry and related fields based on the sections of *Chemical Abstracts. Scientometrics, 78(2), 219-229*

Application I
Leiden Ranking

- Leiden Ranking, first published 2012 (LR 2011/2012) as a bibliometric based research ranking of universities
- Citation data of Web of Science for 500 universities with the largest output (~3.3 Mio publications in the years 2005-2009)
- Main research questions (Bornmann, Mutz & Daniel, 2013):
  1. How to model aggregated citation data?
  2. Are there any real differences in citation impact between universities and countries beyond random fluctuations?
  3. To what extent can such differences be explained by certain covariates?

- Measure: $PP_{top10\%}$, full counting
- Covariates: gross domestic product (GDP (PPP)), number of residents, total area of country, proportion of residents younger than 15 years
- Statistical model: multilevel logistic regression, which considers the hierachical structure of data

Results

Ranking according to the predicted probabilties

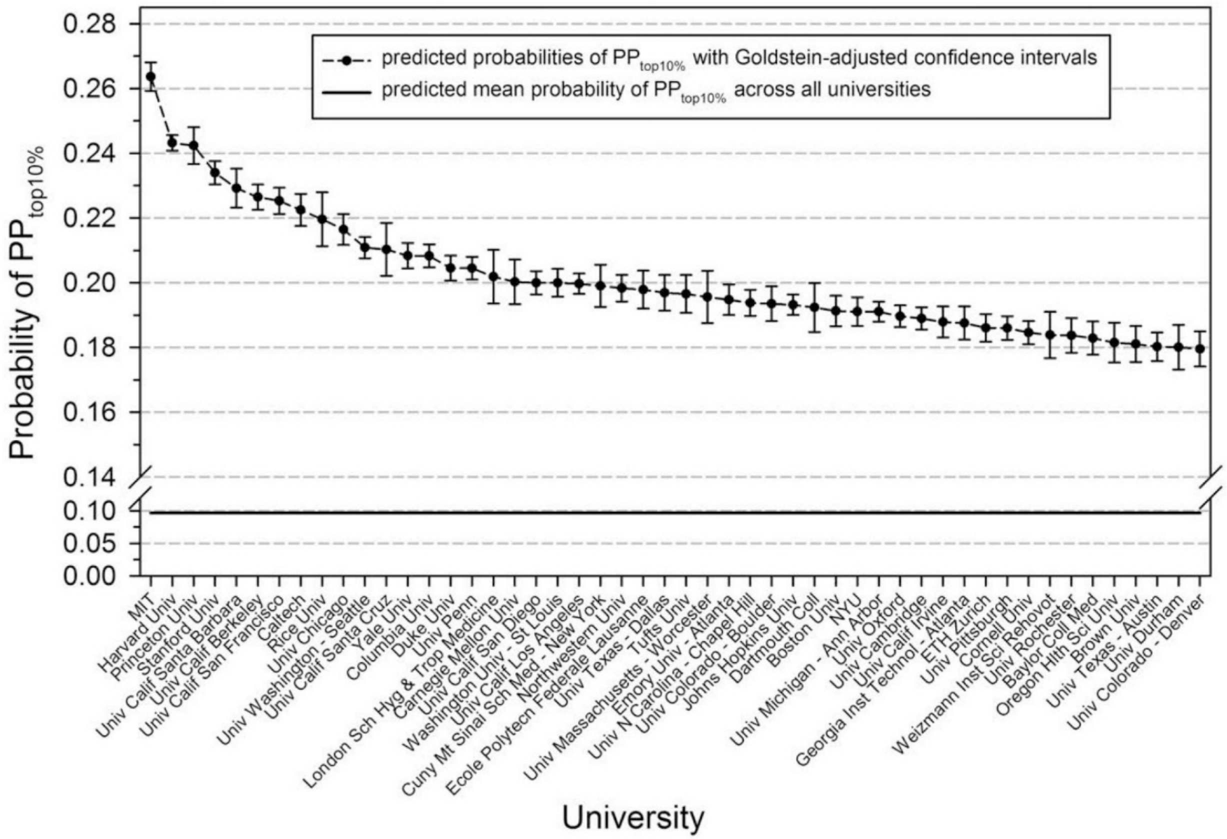


FIG. 1. The 50 highest ranked universities among the total of 500 universities, ranked from left to right according to decreasing $PP_{top10\%}$ probabilities.

5.5% of the $PP_{top10\%}$ variance is attributable to differences between universities.

Ranking adjusted for covariates

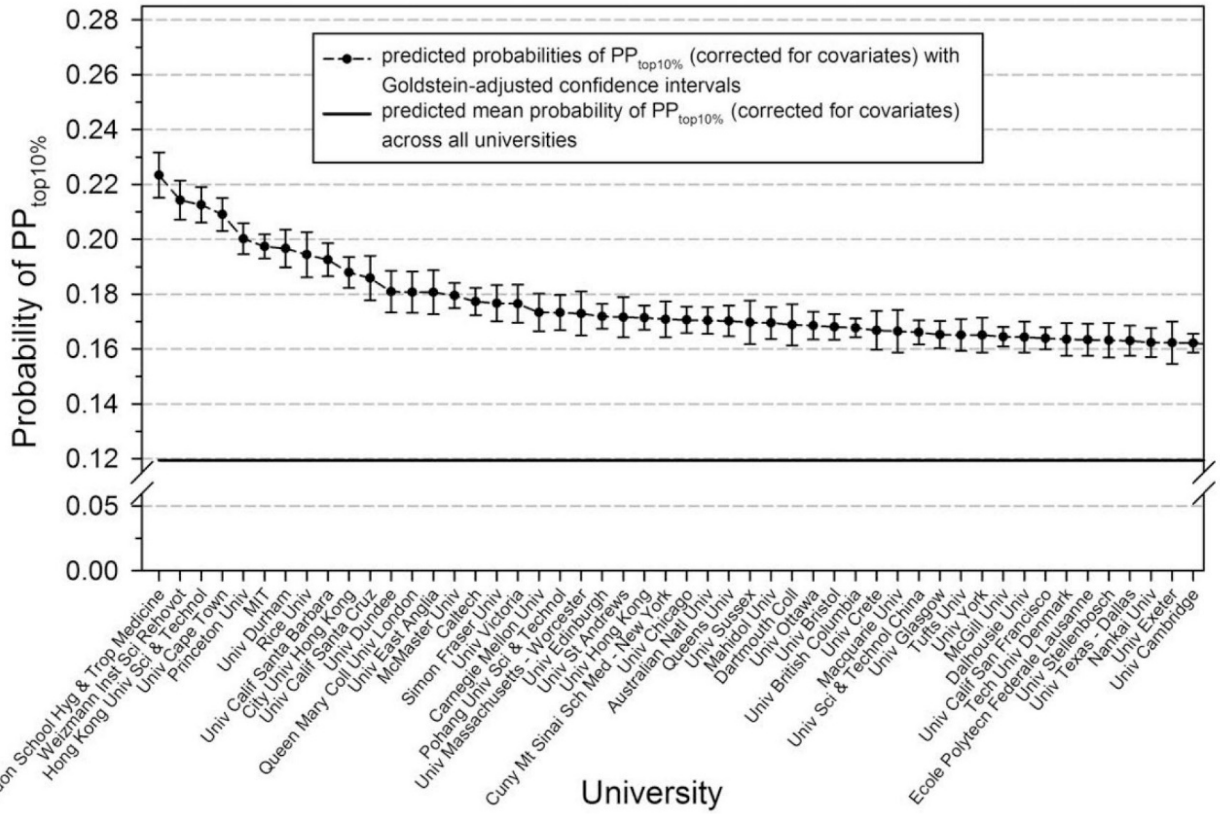


FIG. 3. Covariate-adjusted ranking of the 50 highest ranked universities among the total of 500 universities, ranked from left to right according to decreasing $PP_{top10\%}$ probabilities.

Ranking of countries

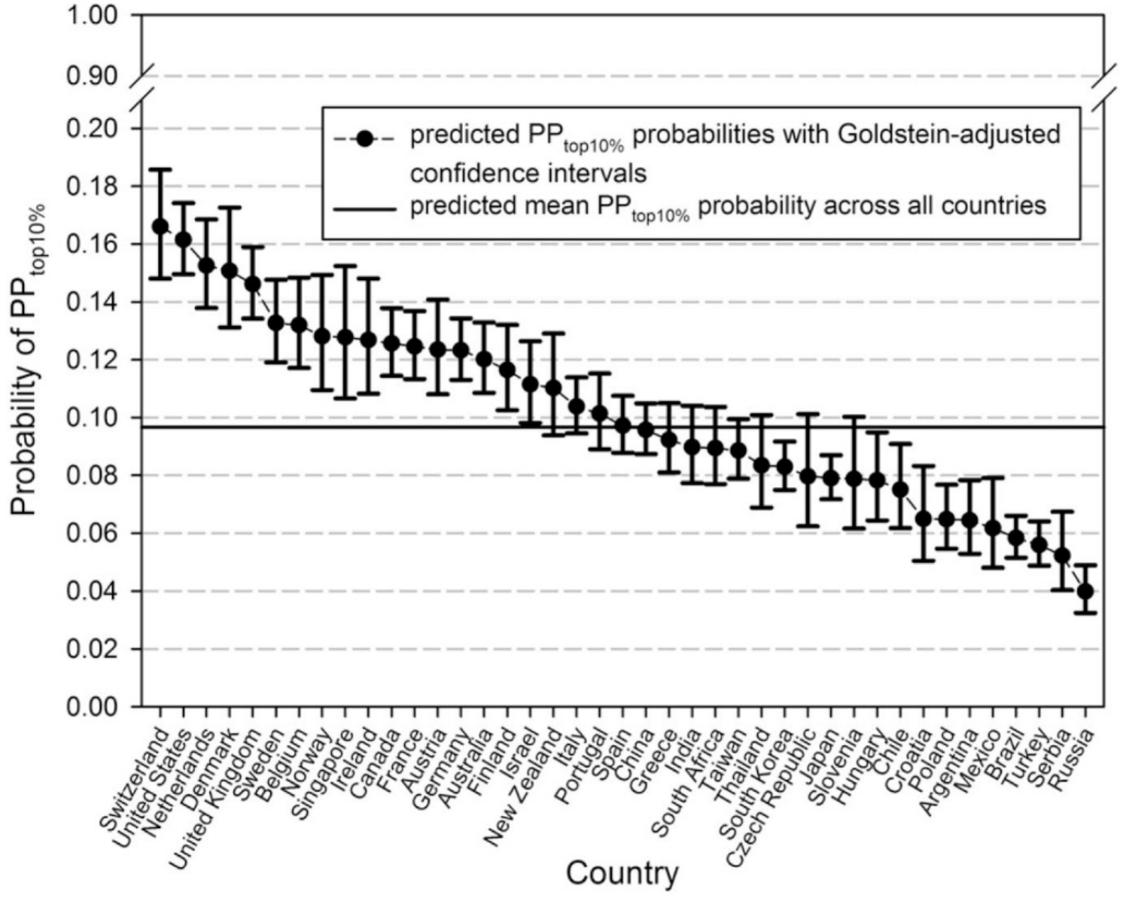


FIG. 2. Ranking of all countries with at least one university in the Leiden Ranking, ranked from left to right according to decreasing $PP_{top10\%}$ probabilities.

78.2% of the systematic variance of $PP_{top10\%}$ between universities can be explained by differences between countries.

Application II
Excellence Mapping

www.excellencemapping.net

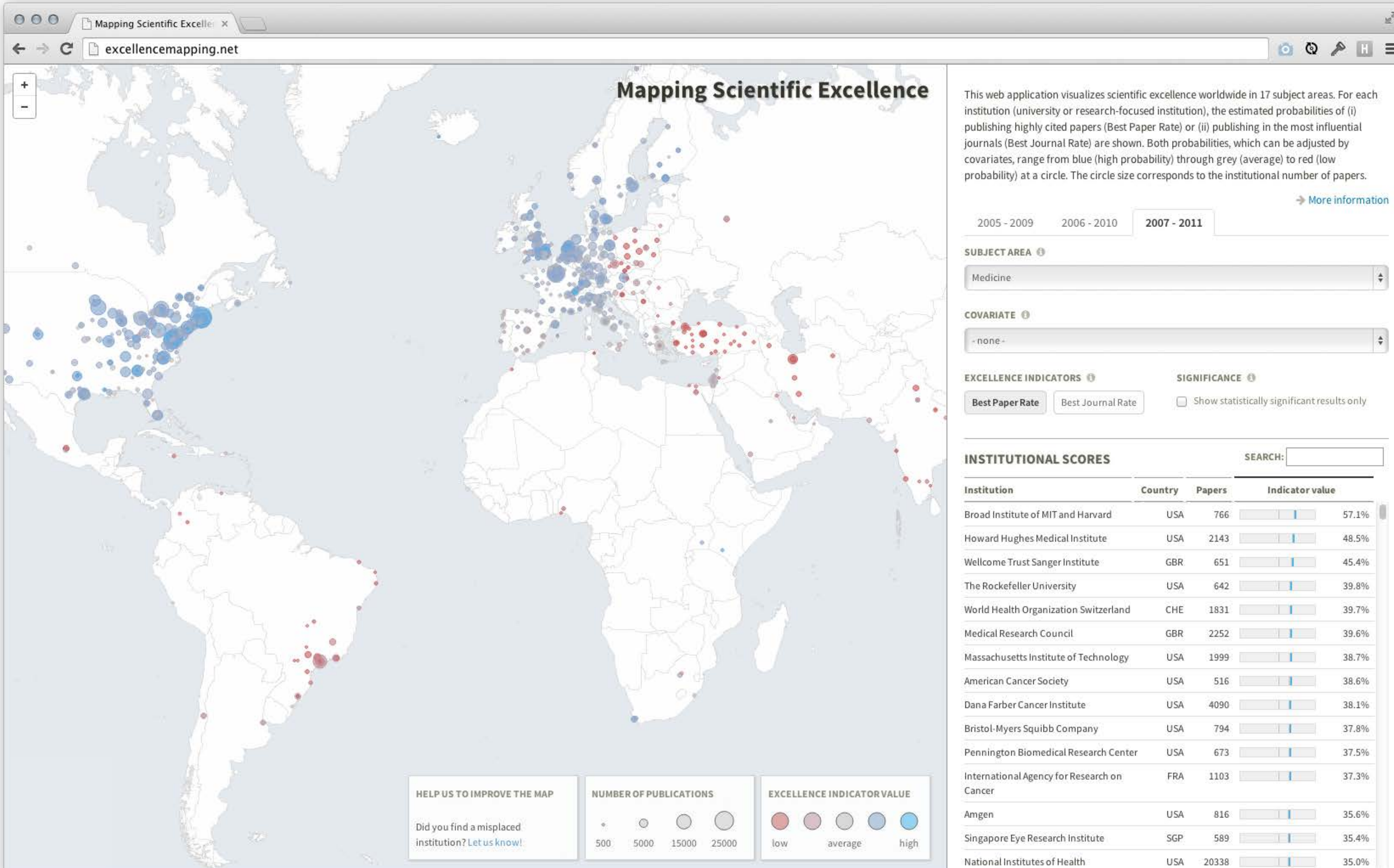
- In recent years, spatial visualization approaches have been introduced in scientometrics
- For example, maps have been published identifying hot regions of scientific performance
- excellencemapping.net combines both approaches
- Institutional performance is presented as a ranking and on a map
- Team: Lutz Bornmann (bibliometrics, Max-Planck Society), Felix de Moya Anegón (data, SCImago), Moritz Stefaner (grafic design), Rüdiger Mutz (statistics, ETHZ)

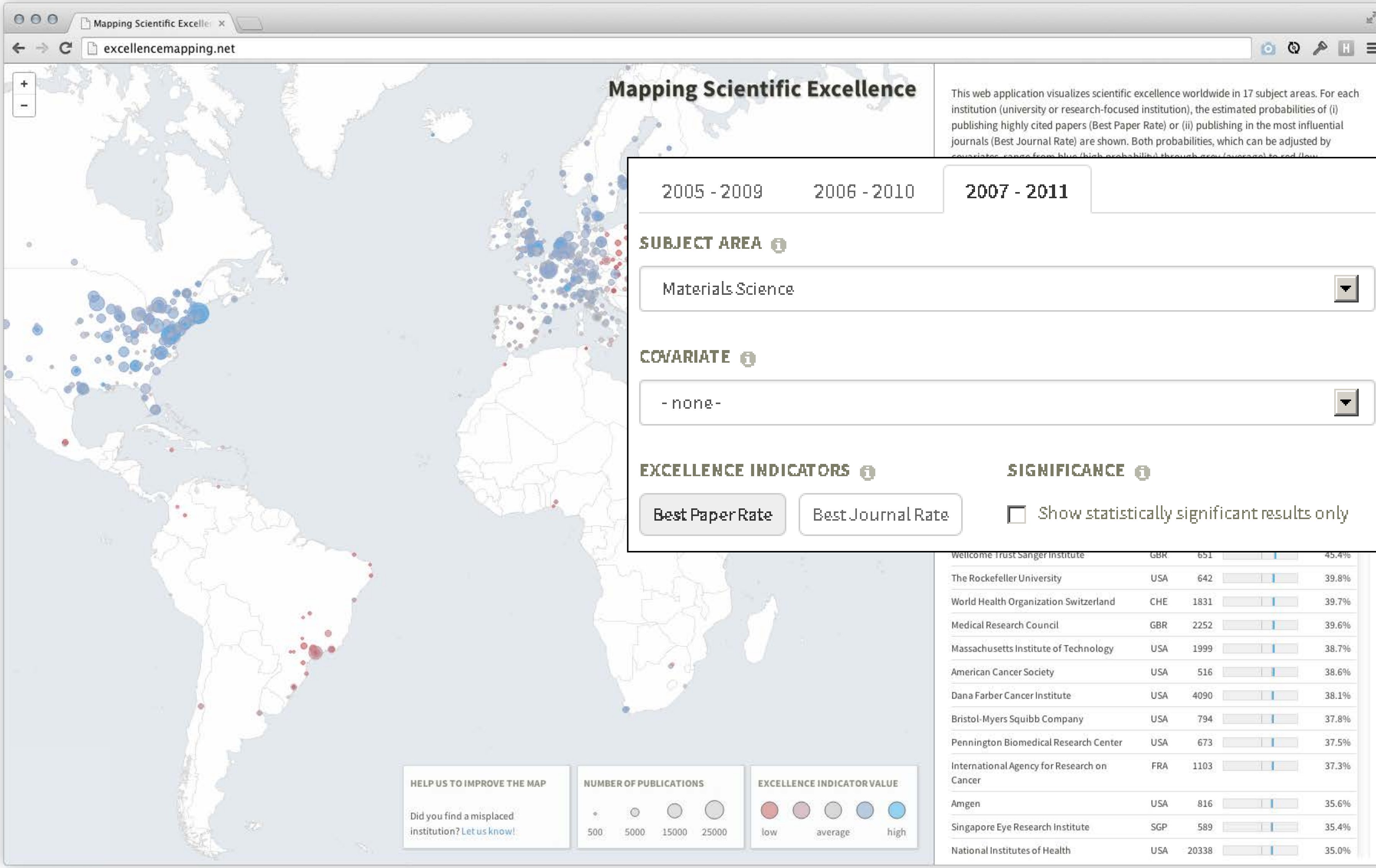
Data

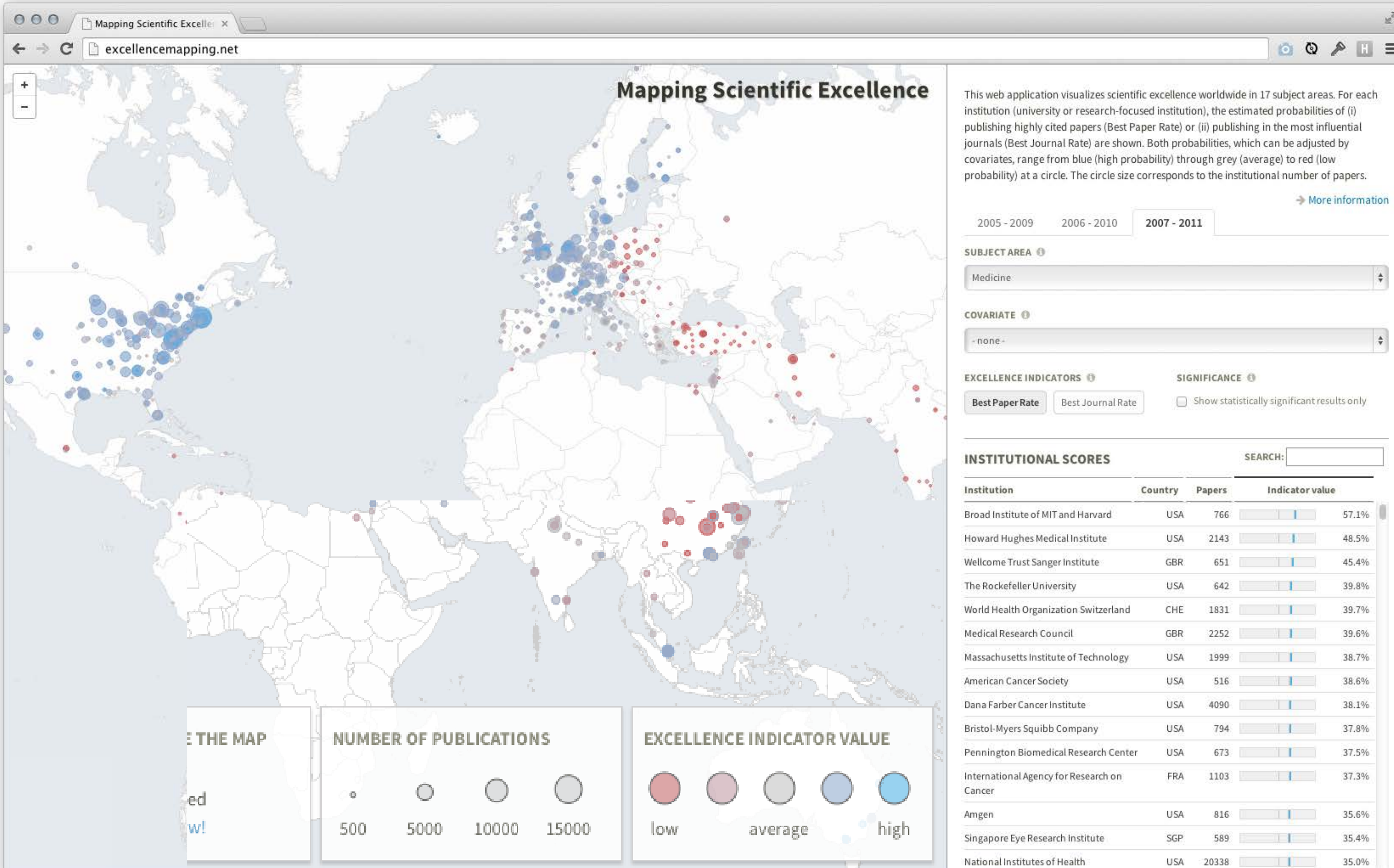
- Scopus data – institutional addresses have been cleaned by SCImago
- Universities and research-focused institutions
- Articles, Reviews und Conference Papers published between 2007 and 2011 within a subject category (third release)
- Only institutions, which have published at least 500 Papers within a subject category
- Full counting: Independent of the number of co-authoring institutions, an institution on a paper receives the full credit
- Indicators measuring performance: best journal rate und best paper rate ($PP_{top10\%}$)
- 17 subject areas (e.g., Chemistry, Neuroscience). Areas with less than 50 institutions are not considered

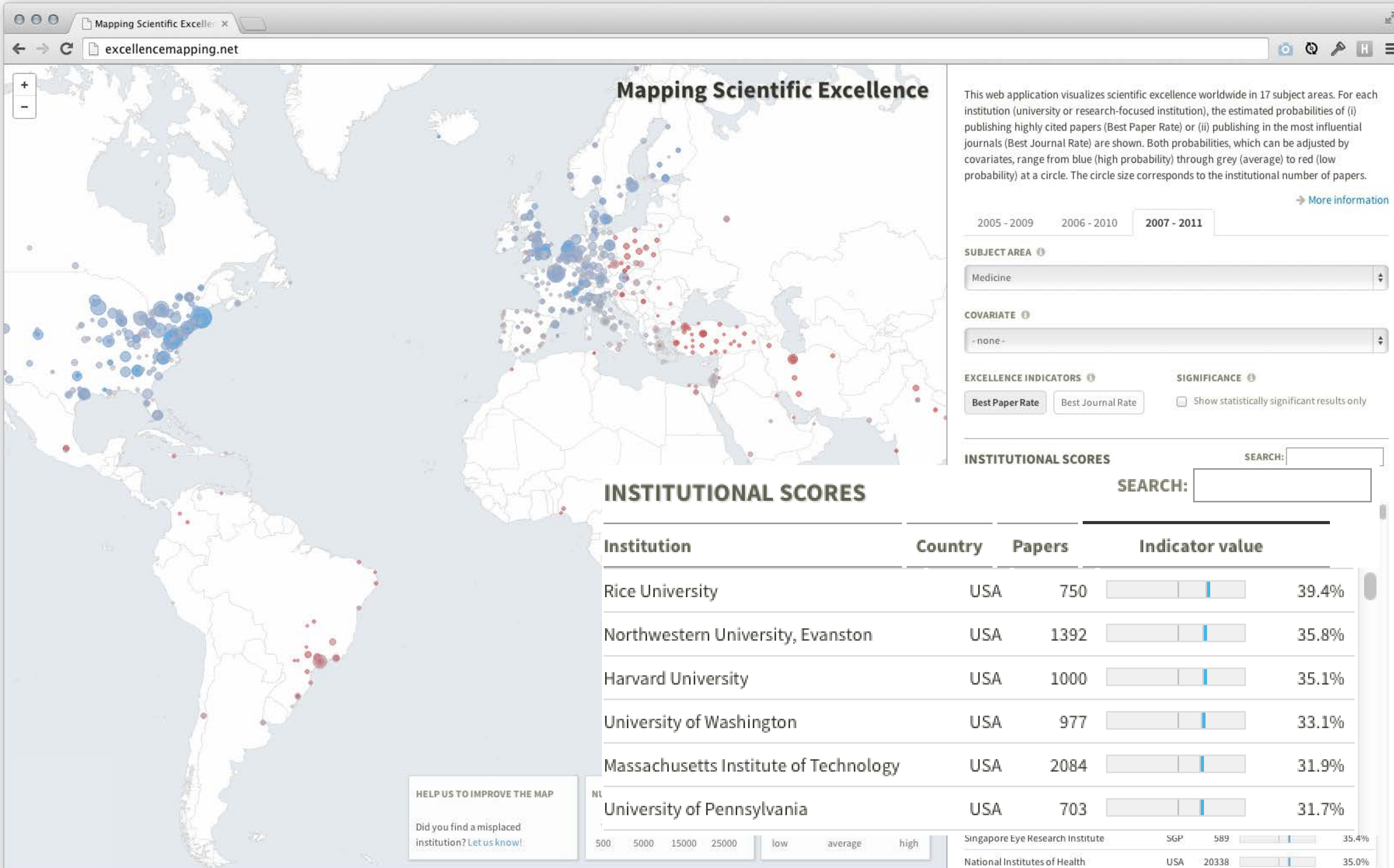
Statistical Model

- excellencemapping.net presents results of multilevel regression models (so called predicted values)
- Dependent variable: performance indicator
- aggregates: all institutions within a subject area
- Further covariates: Factors (e.g., Gross Domestic Product, Corruption Index) with a possible influence on institutional performance


Mapping Scientific Excellence
excellencemapping.net
Mapping Scientific Excellence
This web application visualizes scientific excellence worldwide in 17 subject areas. For each institution (university or research-focused institution), the estimated probabilities of (i) publishing highly cited papers (Best Paper Rate) or (ii) publishing in the most influential journals (Best Journal Rate) are shown. Both probabilities, which can be adjusted by covariates, range from blue (high probability) through grey (average) to red (low probability) at a circle. The circle size corresponds to the institutional number of papers.
More information
2005 - 2009
2006 - 2010
2007 - 2011
SUBJECT AREA
Medicine
COVARIATE
- none -
EXCELLENCE INDICATORS
Best Paper Rate
Best Journal Rate
SIGNIFICANCE
Show statistically significant results only
INSTITUTIONAL SCORES
SEARCH:
Institution Country Papers Indicator value
Broad Institute of MIT and Harvard USA 766 57.1%
Howard Hughes Medical Institute USA 2143 48.5%
Wellcome Trust Sanger Institute GBR 651 45.4%
The Rockefeller University USA 642 39.8%
World Health Organization Switzerland CHE 1831 39.7%
Medical Research Council GBR 2252 39.6%
Massachusetts Institute of Technology USA 1999 38.7%
American Cancer Society USA 516 38.6%
Dana Farber Cancer Institute USA 4090 38.1%
Bristol-Myers Squibb Company USA 794 37.8%
Pennington Biomedical Research Center USA 673 37.5%
International Agency for Research on Cancer FRA 1103 37.3%
Amgen USA 816 35.6%
Singapore Eye Research Institute SGP 589 35.4%
National Institutes of Health USA 20338 35.0%
HELP US TO IMPROVE THE MAP
Did you find a misplaced institution? Let us know!
NUMBER OF PUBLICATIONS
500 5000 15000 25000
EXCELLENCE INDICATOR VALUE
low average high


Mapping Scientific Excelle
excellencemapping.net
Mapping Scientific Excellence
This web application visualizes scientific excellence worldwide in 17 subject areas. For each institution (university or research-focused institution), the estimated probabilities of (i) publishing highly cited papers (Best Paper Rate) or (ii) publishing in the most influential journals (Best Journal Rate) are shown. Both probabilities, which can be adjusted by
2005 - 2009
2006 - 2010
2007 - 2011
SUBJECT AREA
Materials Science
COVARIATE
- none -
EXCELLENCE INDICATORS
Best Paper Rate
Best Journal Rate
SIGNIFICANCE
Show statistically significant results only
Wellcome Trust Sanger Institute GBR 651 45.4%
The Rockefeller University USA 642 39.8%
World Health Organization Switzerland CHE 1831 39.7%
Medical Research Council GBR 2252 39.6%
Massachusetts Institute of Technology USA 1999 38.7%
American Cancer Society USA 516 38.6%
Dana Farber Cancer Institute USA 4090 38.1%
Bristol-Myers Squibb Company USA 794 37.8%
Pennington Biomedical Research Center USA 673 37.5%
International Agency for Research on Cancer FRA 1103 37.3%
Amgen USA 816 35.6%
Singapore Eye Research Institute SGP 589 35.4%
National Institutes of Health USA 20338 35.0%
HELP US TO IMPROVE THE MAP
Did you find a misplaced institution? Let us know!
NUMBER OF PUBLICATIONS
500 5000 15000 25000
EXCELLENCE INDICATOR VALUE
low average high


Mapping Scientific Excelle
excellencemapping.net
Mapping Scientific Excellence
This web application visualizes scientific excellence worldwide in 17 subject areas. For each institution (university or research-focused institution), the estimated probabilities of (i) publishing highly cited papers (Best Paper Rate) or (ii) publishing in the most influential journals (Best Journal Rate) are shown. Both probabilities, which can be adjusted by covariates, range from blue (high probability) through grey (average) to red (low probability) at a circle. The circle size corresponds to the institutional number of papers.
More information
2005 - 2009
2006 - 2010
2007 - 2011
SUBJECT AREA
Medicine
COVARIATE
- none -
EXCELLENCE INDICATORS
SIGNIFICANCE
Best Paper Rate
Best Journal Rate
Show statistically significant results only
INSTITUTIONAL SCORES
SEARCH:
Institution
Country
Papers
Indicator value
Broad Institute of MIT and Harvard USA 766 57.1%
Howard Hughes Medical Institute USA 2143 48.5%
Wellcome Trust Sanger Institute GBR 651 45.4%
The Rockefeller University USA 642 39.8%
World Health Organization Switzerland CHE 1831 39.7%
Medical Research Council GBR 2252 39.6%
Massachusetts Institute of Technology USA 1999 38.7%
American Cancer Society USA 516 38.6%
Dana Farber Cancer Institute USA 4090 38.1%
Bristol-Myers Squibb Company USA 794 37.8%
Pennington Biomedical Research Center USA 673 37.5%
International Agency for Research on Cancer FRA 1103 37.3%
Amgen USA 816 35.6%
Singapore Eye Research Institute SGP 589 35.4%
National Institutes of Health USA 20338 35.0%
E THE MAP
ed
w!
NUMBER OF PUBLICATIONS
500
5000
10000
15000
EXCELLENCE INDICATOR VALUE
low
average
high


Mapping Scientific Excellence
excellencemapping.net
This web application visualizes scientific excellence worldwide in 17 subject areas. For each institution (university or research-focused institution), the estimated probabilities of (i) publishing highly cited papers (Best Paper Rate) or (ii) publishing in the most influential journals (Best Journal Rate) are shown. Both probabilities, which can be adjusted by covariates, range from blue (high probability) through grey (average) to red (low probability) at a circle. The circle size corresponds to the institutional number of papers.
More information
2005 - 2009
2006 - 2010
2007 - 2011
SUBJECT AREA
Medicine
COVARIATE
- none -
EXCELLENCE INDICATORS
Best Paper Rate
Best Journal Rate
SIGNIFICANCE
Show statistically significant results only
INSTITUTIONAL SCORES
SEARCH:
Institution
Country
Papers
Indicator value
Rice University USA 750 39.4%
Northwestern University, Evanston USA 1392 35.8%
Harvard University USA 1000 35.1%
University of Washington USA 977 33.1%
Massachusetts Institute of Technology USA 2084 31.9%
University of Pennsylvania USA 703 31.7%
Singapore Eye Research Institute SGP 589 35.4%
National Institutes of Health USA 20338 35.0%
HELP US TO IMPROVE THE MAP
Did you find a misplaced institution? Let us know!
500 5000 15000 25000
low average high

- In 2011, first release of excellencemapping.net
- Since then a lot of feedback (MIT Technology Review)
- Last year, the third release has been published
- Mid 2014 39,000 users
- Next step: Excellence-Network to represent inter-institutional collaborations (~Mid 2015)

- Limitations:
  - only two indicators are used for measuring research performance
  - citations measure impact and not quality (impact is one part)
  - data problems: erroneous addresses, address on paper is not the location of research, wrong geocodes, …

Conclusions

- A numerical or statistical perspective on bibliometric data have serious consequences on the kind of processing of data, the analysis of data and the final interpretation of the aggregated results (e.g., rankings).
- Classical bibliometric indicators as Crown indicators or any "transformations" of data (e.g., fractional counting) are difficult to manage in statistical analyses. -> preference for raw data
- Inter-institutional comparisons of universities in their research performance require some adjustment for bias factors as GDP.
- Field specific rankings are preferred towards global ranking across all fields.

„… the personal wish of the author remains to send all bibliometrics and its diligent servants to the darkest omnivoric black hole that is known in the entire universe, in order to liberate academia forever from this pestilence.“

Richard R. Ernst

(ETHZ, Nobel prize for chemistry, 1991)

Ernst, Richard R. (2010). The follies of citation indices and academic ranking lists a brief commentary to „Bibliometrics as Weapons of Mass Citation“. *CHIMIA,64*(1/2), p. 90.

Many thanks for your attention!