How researchers' performance should be assessed: Wrap up

Yerevan 27-28 Oct. 2022

Giovanni Abramo
Laboratory for Studies in Research Evaluation
Institute for System Analysis and Computer Science
National Research Council of Italy

Outline

1. Research evaluation: why
2. How to measure (and compare!) individual research performance
3. How not to measure research performance
4. Ranking distortions when using invalid indicators
5. Conclusions and recommendations

Conclusions

- ✓ Count only what counts and be aware of what you cannot count
- ✓ The most popular research performance indicators are invalid
- ✓ Field classification of scientists is absolutely required to compare performance at the individual level
- ✓ Research performance at the individual level is absolutely required to measure performance at organizational level

The use of bibliometrics to inform decision-making at policy and management levels

Yerevan 27-28 Oct. 2022

Giovanni Abramo
Laboratory for Studies in Research Evaluation
Institute for System Analysis and Computer Science
National Research Council of Italy

Research evaluation goals

- ✓ Stimulating higher research productivity
- ✓ Allocating resources according to performance
- ✓ Informing research policy (strategy)
- ✓ Reducing information asymmetry between supply and demand
- ✓ Demonstrating that investment in research is effective and delivers public benefits

Outline

1. How to measure research performance at the aggregate level (discipline, institution, …)
2. How not to measure research performance
3. Ranking distortions when using invalid indicators
4. Conclusions and recommendations

Tangible image

"Scientometrics is to research policy/management as diagnostic imaging is to medicine"

Research evaluation problems

- ✓ Proliferation of (invalid) performance indicators
- ✓ Doubtful assessment methods
- ✓ Abundance of non theory-based rankings
- ✓ Media fanfare for (wrong!) world institutions rankings
- ✓ Do-it-yourself practices
- ✓ Poor strategic and policy perspectives

Individual research performance indicator

The Fractional Scientific Strength (FSS)

$$FSS_R = \frac{1}{(w_R + k)} \cdot \frac{1}{t} \sum_{i=1}^{N} c_i \, f_i$$

Where:

N = number of publications of the researcher in the period under observation

c_i = weighted combination of normalized citations and impact factor associated to publication i*

f_i = fractional contribution of the researcher to publication i

w_R = average yearly salary of the researcher

k = average yearly capital used by the researcher

t = number of years of work of the researcher in the period under observation

* Abramo, G., D'Angelo, C.A., & Felici, G. (2019). Predicting long-term publication impact through a combination of early citations and journal impact factor. *Journal of Informetrics*, 13(1), 32-49.

FSS in words

- ✓ It counts all publications of a scientist in a period of time
- ✓ It divides each publication by the number of authors (it weights their contribution based on their position in the byline, if not alphabetically ordered)
- ✓ It measures their value by a weighted combination of citations and IF, each scaled by a factor accounting for field and year of publication
- ✓ It accounts for input

Research performance at the aggregate level

- ✓ Premise: To measure research institution performance, one needs to know the identity of staff, output and field of research of each individual
- ✓ Statement: Research institutions are not homogenous in terms of number and size of research fields
- ✓ Proposition: Individual performance is absolutely required to measure performance at organizational level

The Fractional Scientific Strength (FSS)

aggregate level

Productivity of research units (e.g. field, discipline, department, institution) based on FSS_R

$$FSS_D = \frac{1}{RS} \sum_{j=1}^{RS} \frac{FSS_{R_j}}{\overline{FSS_R}}$$

Where:

RS = research staff of the unit, in the observed period

FSS_{Rj} = productivity of researcher *j* in the research unit

$\overline{FSS_R}$ = average productivity of all national productive researchers in the same SDS as researcher *j*

The performance in each field (SDS)

The fields within the discipline (UDA) "Medicine" of Institution *X*

	O			FSS		
Field	Score	Rank	Rank(η)	Score	Rank	Rank(η)
MED/09-Internal medicine	0.739	6 out of 12	55	0.435	8 out of 12	36
BIO/14-Pharmacology	0.457	25 out of 37	33	0.287	25 out of 37	33
MED/38-General and specialised pediatrics	0.524	33 out of 42	22	0.460	28 out of 42	34
MED/40-Gynaecology and obstetrics	0.816	5 out of 22	81	0.242	5 out of 22	81
MED/42-General and applied hygiene	1.103	8 out of 52	86	1.000	14 out of 52	75
MED/07-Microbiology and clinical microbiology	1.525	3 out of 23	91	2.077	2 out of 23	95
BIO/13-Applied biology	0.425	37 out of 42	12	0.225	38 out of 42	10
MED/08-Pathological anatomy	0.667	28 out of 40	31	0.485	26 out of 40	36
MED/11-Cardiovascular diseases	1.023	8 out of 27	73	1.053	9 out of 27	69

The performance of Institution '*X*' in each discipline (UDA)

	O		SS		FO		FSS	
UDA*	Score	Rank(η)	Score	Rank(η)	Score	Rank(η)	Score	Rank(η)
2	1.231	81	1.246	76	0.988	69	1.087	76
3	1.031	72	0.973	63	1.111	86	1.092	79
5	1.031	65	0.853	45	1.064	72	0.865	47
6	1.033	74	1.033	67	1.080	76	1.115	79
7	0.775	46	0.643	39	0.845	54	0.734	46
9	0.741	26	0.763	43	0.763	33	0.766	39

** 2, Physics; 3, Chemistry; 5, Biology; 6, Medicine; 7, Agricultural and veterinary sciences; 9, Industrial and information engineering*

Key performance indicators

- Individual level:
  - Productivity (FSS) and its components
  - Highly-cited articles per researcher
- Institution level:
  - Productivity (FSS)
  - Share of unproductive staff
  - Share of top scientists
  - Effectiveness of recruitment
  - Rate of institution-industry collaborations

How not to measure research performance

- ✓ The MNCS
- ✓ The h-index
- ✓ The Shanghai ranking and the like
- ✓ The UK-like national research assessment exercises (RAEs)

Distortion of universities rankings by h and g indexes

UDA	Percentage of Q1 universities by FSS not included in the same set by h	Percentage of Q1 universities by FSS not included in the same set by g
Mathematics and computer science	45	47
Physics	48	51
Chemistry	49	46
Earth sciences	42	35
Biology	42	36
Medicine	40	35
Agricultural and veterinary science	41	33
Civil engineering	28	26
Industrial and information engineering	40	35
Total	*42*	*38*

The Shanghai Ranking

ARWU
by Shanghai Jiao Tong University

http://www.shanghairanking.com/ARWU2018.html

Sapienza, Padua: 151-200

Academic Ranking of World Universities 2018

World Rank	Institution*	By location: All	National/Regional Rank	Total Score	Score on: Alumni
1	Harvard University		1	100	100
2	Stanford University		2	75.6	44.5
3	University of Cambridge		1	71.8	82.3
4	Massachusetts Institute of Technology (MIT)		3	69.9	70.9
5	University of California, Berkeley		4	68.3	65.6
6	Princeton University		5	61	55.8
7	University of Oxford		2	60	50.8
8	Columbia University		6	58.2	62.8
9	California Institute of Technology		7	57.4	53.5
10	University of Chicago		8	55.5	59.2
11	University of California, Los Angeles		9	51.2	29.2
12	Cornell University		10	50.7	43.1
12	Yale University		10	50.7	47.1
14	University of Washington		12	50	24.9
15	University of California, San Diego		13	47.8	19

The Shanghai ranking criteria

Metodology: total score

Criteria	Indicator	Weight
Quality of Education	Alumni of an institution winning Nobel Prizes and Fields Medals	10%
Quality of Faculty	Staff of an institution winning Nobel Prizes and Fields Medals	20%
	Highly cited researchers in 21 broad subject categories	20%
Research Output	Papers published in Nature and Science	20%
	Papers indexed in SCI-E and SSCI (Web of Science)	20%
Per Capita Performance	Per capita academic performance of an institution	10%

90% of the score is size dependent!

UK like RAEs: main limits

- ✓ *Robustness:* How sensitive are rankings to the share of the output evaluated?
- ✓ *Reliability:* Do universities submit their best outputs?
- ✓ *Accuracy:* How accurate is the quality evaluation of products and institutions?
- ✓ *Functionality:* How useful are national rankings for universities, students, companies, …?
- ✓ *Costs and time of execution:* Very high

Comparison of Italian RAE and FSS quartile university ranking lists

UDA	No. of universities	% shifting quartile	Correlat.	From top to non top
Mathematics and computer science	50	46.0%	0.60	46.2%
Physics	43	60.5%	0.25	38.5%
Chemistry	42	59.5%	0.69	45.5%
Earth sciences	30	60.0%	0.52	37.5%
Biology	50	52.0%	0.60	69.2%
Medicine	43	48.8%	0.73	45.5%
Agricultural and veterinary sciences	28	46.4%	0.77	42.9%
Industrial and information engineering	46	47.8%	0.56	50.0%

Insight

'Non quia difficilia sunt non audemus, sed quia non audemus difficilia sunt'

Seneca (4 BC - 65 AD)

'*It is not because things are difficult that we do not try; it is because we do not try that things are difficult'*

Conclusions

- ✓ Count only what counts and be aware of what you cannot count
- ✓ The most popular research performance indicators are invalid
- ✓ Field classification of scientists is absolutely required to compare performance at the individual level
- ✓ Research performance at the individual level is absolutely required to measure performance at organizational level
- ✓ Avoid the “do-it-yourself” temptation
- ✓ The FSS performance evaluation conducted in Italy can be replicated in any other institution or country, and the first might serve as a benchmark for comparison

Grazie

Շնորհակալություն

Thank you!

Giovanni Abramo

Laboratory for Studies in Research Evaluation

at the Institute for System Analysis and Computer Science (IASI-CNR)

National Research Council of Italy

Via dei Taurini 19, 00185 Roma – ITALY
Tel. +39 06 72597362

giovanni.abramo@uniroma2.it

https://sites.google.com/view/laboratoriortt