

# **Metrics-Based Assessment of Scientific-Scholarly Research**

**Henk F. Moed**

**Lecture at the seminar ‘How to Measure  
Estonian Productivity?’**

**Tartu, 6 Nov 2014**

# Short CV Henk F. Moed

1981-2009	Staff member at Centre for Science and Technology Studies ( <b>CWTS</b> ), Leiden Univ.
2009	<b>Professor</b> of Research Assessment Methodologies at Leiden University
2010 – Sept 2012	<b>Elsevier</b> , SciVal Dept. Senior Scientific <b>Advisor</b>
Sept 2012 – 1 Nov 2014	Elsevier, AGRM Dept. Head of <b>Informetric Research Group</b>
Sept-Oct 2014	<b>Visiting professor</b> at Univ Rome la Sapienza ( <b>SSSA</b> )

# 10 Main Issues

- 1 Combine metrics and peer review**
- 2 Design of assessment process depends upon context**
- 3 Performance-based funding can be highly targeted**
- 4 Popular journal impact factors are often misinterpreted**
- 5 Role of journal impact factors must be reduced**
- 6 More advanced indicators are technically feasible**
- 7 Advanced online self-assessment tools are needed**
- 8 Manuscript peer review is essential**
- 9 Focus not merely on output, also on input and process**
- 10 Metrics should be independent and have wide access**

# 1. Combine metrics and peer review

- Let the type of peer review depend upon the outcomes of a bibliometric study
- Use peer review to assess cases in which metrics indicate extreme scores (“outliers”)

## 5 combinations of 'metrics' and peer review

- Policy makers may let the **type** of peer review depend upon the outcomes of a bibliometric study
- Peer committees may use citation analysis for initial rankings and explicitly justify **why** their judgments deviate
- Use peer review if **no valid**, useful, generally accepted metrics are available
- Use peer review to assess cases in which metrics indicate **extreme scores** ("outliers")
- Use peer review if information from **different sources** has to be integrated into a final judgment

# Cicchetti (1991) conclusions on agreement among referees

Evaluation Object	Subject field	Agreement among reviewers
Journal manuscripts	General and diffuse disciplines	Reviewers (and editors) tend to <b>agree more on rejection</b> than on acceptance of manuscript
	Specialized and focused disciplines	Reviewers <b>agree more on acceptance</b> than on rejection
Grant proposals	Science (NSF proposals)	Reviewers <b>agree much more about what is unworthy</b> of support than about what does have scientific value

## Correlations between peer ratings and citation impact of 307 PHYS, CHEM, BIOL research groups (NL-VSNU)

- Citation impact discriminated **very well** between groups rated **excellent or good** and those receiving **lower** peer ratings
- But it did **not** discriminate well between **good** and **excellent** groups in the perception of the peers

## **2. Design of an assessment process depends upon the context**

- **Multi-dimensional Research Assessment Matrix (EC AUBR Expert Group)**
- **Depends upon unit, aspect and objective of the assessment, and on “meta” assumptions**



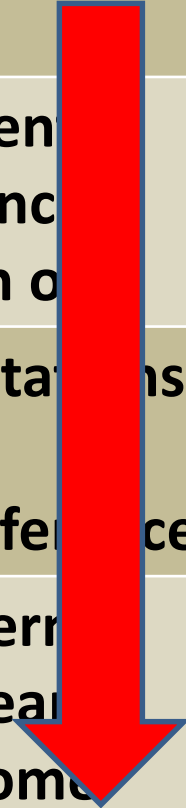
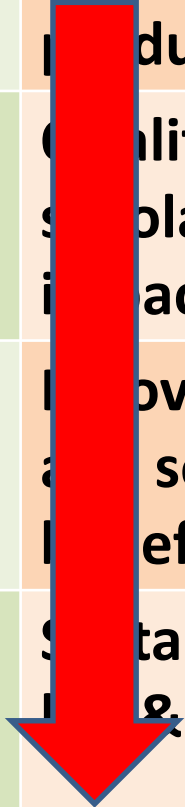
# Multi-dimensional Research Assessment Matrix (Part)

Unit of assessment	Purpose	Output dimensions	Bibliometric indicators	Other indicators
Individual	Allocate resources	Research productivity	Publications	Peer review
Research group	Improve performance	Quality, scholarly impact	Journal citation impact	Patents, licences, spin offs
Department	Increase multi-discipl. research	Innovation and social benefit	Actual citation impact	Invitations for conferences
Institution	Increase regional engagement	Sustainability & Scale	Internat. co-authorship	External research income
Research field	Promotion, hiring	Research infrastruct.	citation 'prestige'	PhD completion rates

# Multi-dimensional Research Assessment Matrix (Part)

Unit of assessment	Purpose	Output dimensions	Bibliometric indicators	Other indicators
Individual	Allocate resources	Research products	Publications	Peer review
Research group	Improve performance	Quality, scholarly impact		en... nc... n o...
Department	Increase multi-discipl. research	Innovat... a... soc... benefit		ita... ns fer... ces
Institution	Increase regional engagement	Sustain... I... & S...		err... ear... ome...
Research field	Promotion, hiring	Research infrastruct.	'prestige'	D com- pletion rates

**Read  
column-  
wise**



# Multi-dimensional Research Assessment Matrix (Part)

Unit of assessment	Purpose	Output dimensions	Bibliometric indicators	Other indicators
Individual	Allocate resources	Research productivity	Publications	Peer review
Research group	Improve performance	Quality, scholarly impact	Journal citation impact	Patents, licences, spin offs
Department	Increase multi-discipl. research	Innovation and social benefit	Actual citation impact	Invitations for conferences
Institution	Increase regional engagement	Sustainability & Scale	Internat. co-authorship	External research income
Research field	Promotion, hiring	Research infrastruct.	citation 'prestige'	PhD completion rates

**Indicators that are appropriate in one context may be useless or invalid in another** [Moed & Halevi, JASIST, 2014a]

**The choice of indicators depends upon:**

**1 Which type of units are to be assessed?**

**2 Which aspects are being assessed?**

**3 What are the assessment's objectives?**

**4 What is the state of the system under assessment? (Meta-analysis)**

**1**

# Aggregation levels / units of analysis

**1 Individual article**

**2 Individual researcher**

**3 Research group**

**4 Research department**

**5 Research institution or center**

**6 Research network**

**7 Research program**

**8 Geographical region (city agglomeration, province)**

**9 Country**

**10 World region**

# Individual researcher

- **PRO**
- **Metrics reveal differences in impact between individuals**

- **CON**
- **Most research articles are the result of team work and are multi-authored**
- **How do we then assess the role of an individual in a team?**

**2****Main types of research output and impact**

<b>Impact</b>	<b>Publication/text</b>	<b>Non-publication</b>
<b>Scientific-scholarly</b>	<b>Scientific journal paper; chapter; monograph</b>	<b>Research data file; video of experiment</b>
<b>Educational</b>	<b>Teaching course book; syllabus; text, hand book</b>	<b>Skilled researchers (e.g., doctorates)</b>
<b>Economic or technological</b>	<b>Patent; commissioned research report</b>	<b>Product; process; spin off; device; design;</b>
<b>Other societal or cultural</b>	<b>Medical guidelines; news-paper article; social media comm.</b>	<b>Interviews; events; performances; advisory work;</b>

Objective	Comments; examples
Allocate resources	Main objective in many national assessments
Improve performance	Base idea: reward (e.g., fund, make visible) performance
Promotion, hiring	At level of individuals; DORA against use journal impact factors
Increase regional engagement	Societal, economic, technological impact (e.g, EC)
Support young researchers	Te be discussed later



## Typical examples

Substantial part of professors in this country is **not research-active** (e.g., too much engaged in teaching)

Through **self-selection** applicants are research active; their quality level tend to be **high**

Researchers in this country are **not well integrated** into the international community; publish mainly in **national** journals

**Young research groups** have **no** good chances to develop in this funding system

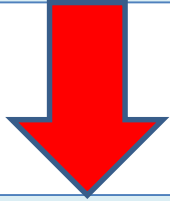
Decisions on hiring tend to be made on the basis of **political considerations**

# Meta analysis

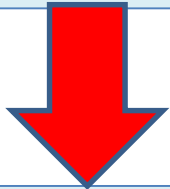
- Important role for policy research and **biblio/informetrics**
- Does **not** aim to assess **individual** units
- Focuses on the **system as a whole**
- Provides a **background** for properly **designing** an assessment process

# CASE 1 [My view: non-defensible use]

**Meta level :  
Policy issue**

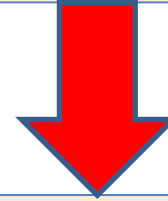


**Policy measure**

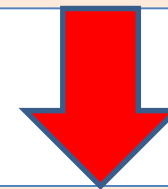


**Bibliometric  
operationali-  
zation**

**Recruitment of the best researchers  
at research universities**



**Use metrics to identify the best  
researchers**



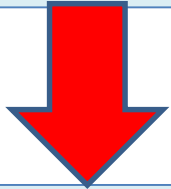
**Rank researchers by average  
impact factor of journals in which  
they published and select nr. 1**

## **CASE 2 [My view: defensible use]**

**Meta level :  
Policy issue**

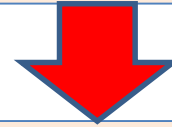


**Policy measure**

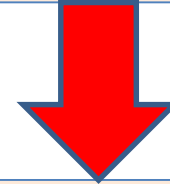


**Bibliometric  
operationali-  
zation**

**Research community is not  
sufficiently oriented toward  
international networks**



**Stimulate publication in good  
international journals**



**Count and reward articles in the  
first impact quartile of journals in  
subject field**

## **CASE 3 [My view: Defensible use]**

**Meta level :  
Policy issue**



**Policy measure**



**Bibliometric  
operationali-  
zation**

**Professors are legally bound to do research but many of them are not research active**



**Allow only research active professors to decide on recruitment of new research staff**

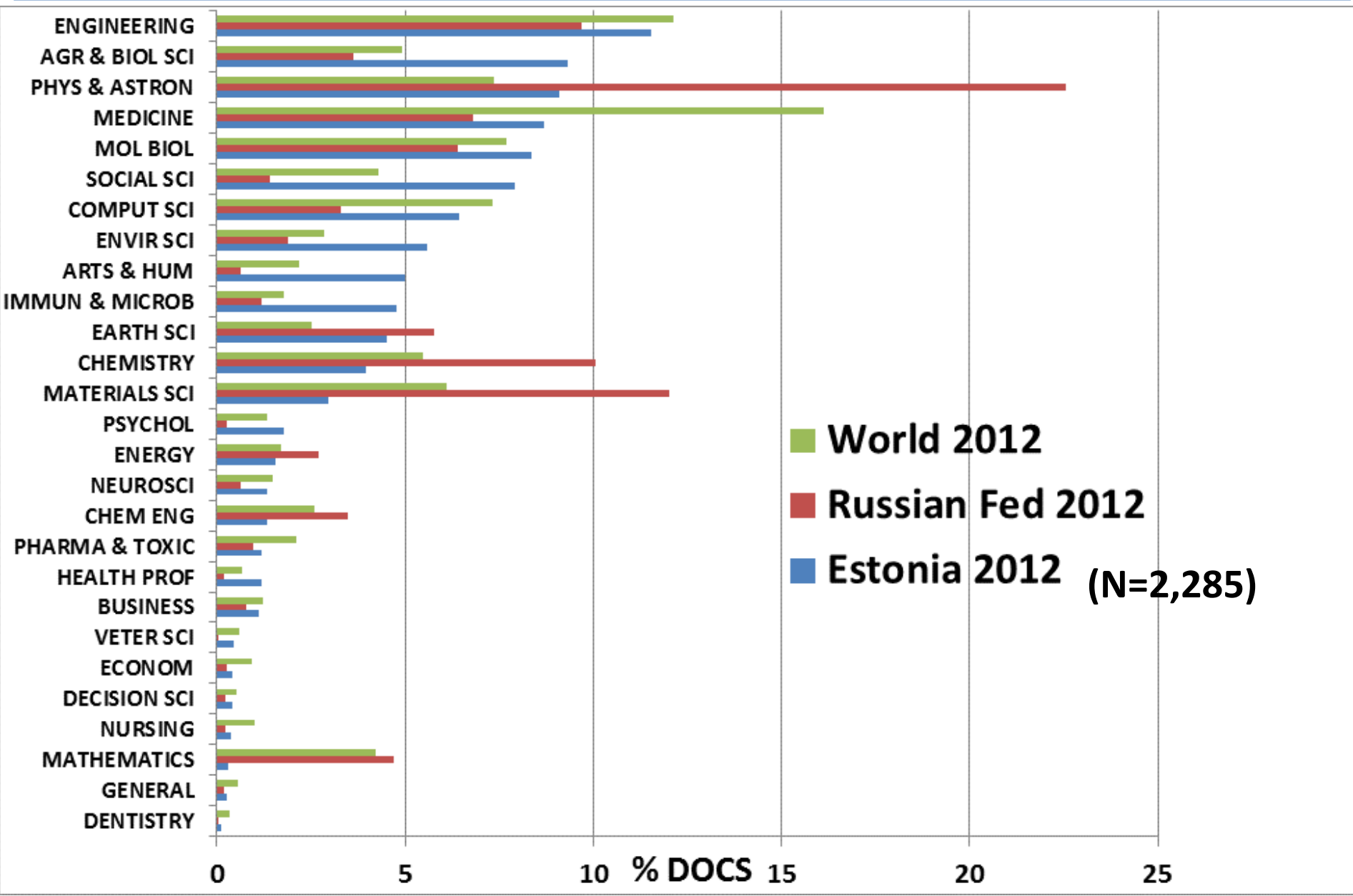


**Select only professors with  $\geq 3$  publications in 7 years in recruitment committees**

# Wider issues

- **Change** an assessment method every **5-10** years?
- Focus on **top** or on **bottom** of quality distribution?
- What is an acceptable **“error rate”**?
- Wrong in **individual** cases  $\Leftrightarrow$  beneficiary for the system **as a whole**

# Disciplinary profile Estonia compared to Russia and world



## Countries with disciplinary profiles similar to Estonia

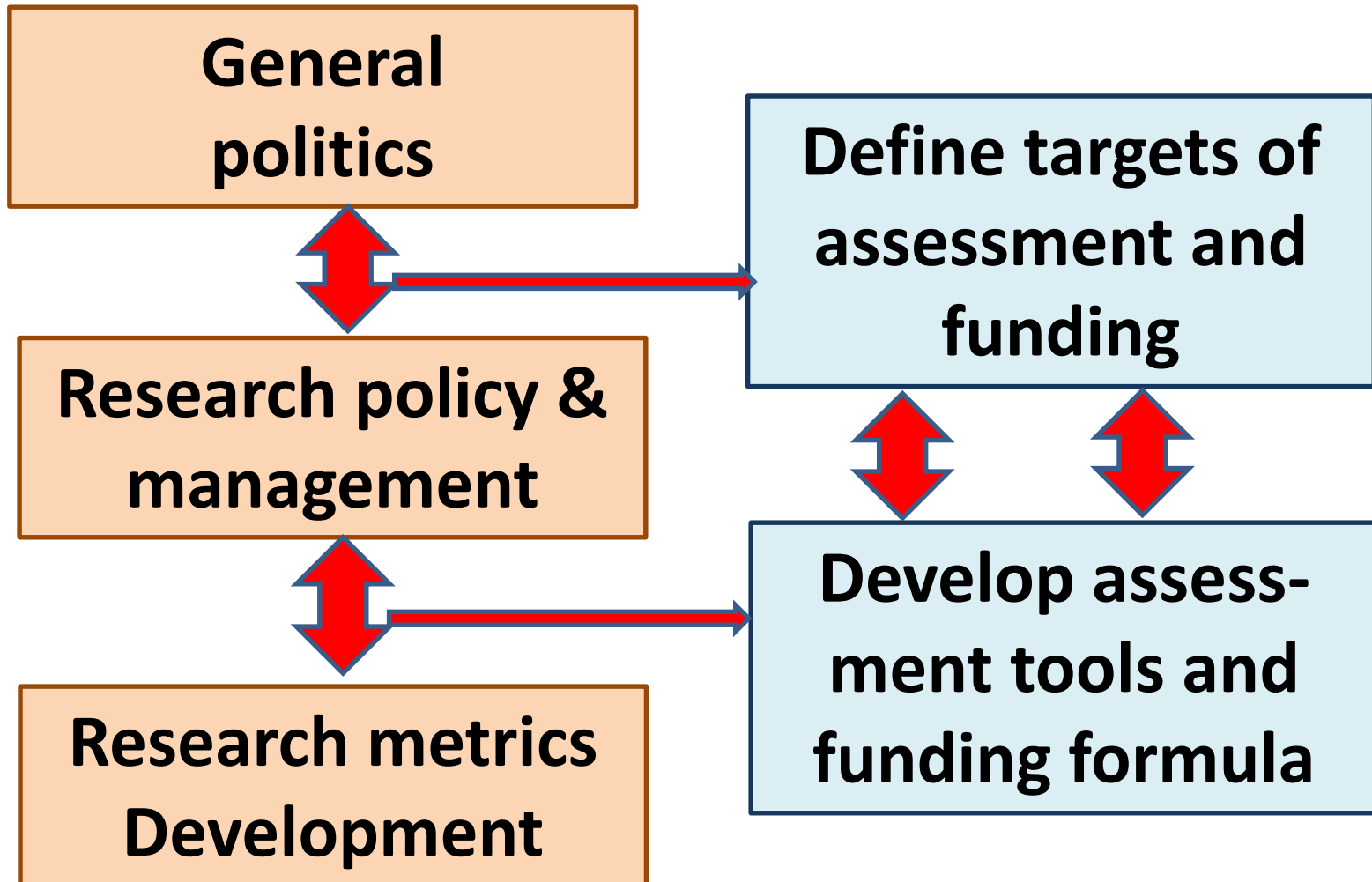
Country	Similarity /Overlap
Mexico	0.90
Slovakia	0.87
Portugal	0.86
Indonesia	0.85
Chile	0.84
Czech Republic	0.83
Slovenia	0.83
India	0.82
Argentina	0.81
Poland	0.81
Finland	0.80
Canada	0.79
Hungary	0.79
Lithoania	0.78
France	0.77



### **3. Performance-based funding can be made highly targeted**

- **Performance based funding could focus on promising emerging research groups**
- **This is feasible without large scale national assessments covering all academic research.**

# Securing a political basis for academic research



# Assessment targets:

- Identifying promising **emerging research groups**
- Identifying **declining research departments**
- Identifying **research-inactive staff members**

# A proposal for targeted assessment and funding

**Meta level:  
Policy objective**

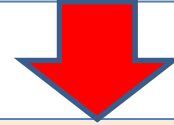


**Policy measure:  
Targeted assess-  
ment & funding**



**Operationali-  
zations**

**Stimulate research quality and the  
introduction of new ideas in the  
academic research system**



- Reward emerging groups;
- Discourage declining depts and non research active staff



- Develop for each target an appropriate set of indicators;
- Develop fit-for-purpose funding formula

# Identify emerging groups (5 yrs)

- **Features**
- **Young senior**
- **Successful PhD and post-doc stage**
- **Expanding competitive funding**
- **Increasing publ output; emerging citat. impact**
- **Growing prestige**

- **Indicators**
- **Age**
- **Status senior's prev.inst citations to PhD oeuvre**
- **Trend in competitive funding**
- **Relative citation rate; diversity journal packet**
- **Speaker invitations; internat. collaboration**

# Proposed process

1. **Define the criteria** for 'emerging group'
2. **Institutions** submit information on their emerging groups to an evaluation unit
3. Data are **validated** and **bibliometric data are added**
4. **International peer review committees** evaluate the submissions
5. **A part of** public funding (block grant) is allocated to institutions (lump sum) on the basis of the number of **emerging groups**

# Proposal: pros and cons

- **PRO**
- **Focuses on groups rather than individuals**
- **Focus on emerging groups stimulates innovation**
- **Uses multiple indicators, combining metrics + peer review**
- **No full scale assessment needed**

- **CON**
- **Difficulties in identifying genuine groups**
- **What to do in social sci and humanities**
- **Does not take into account declining departments or research-inactive staff**

## **4. Journal impact factors are popular but often misinterpreted**

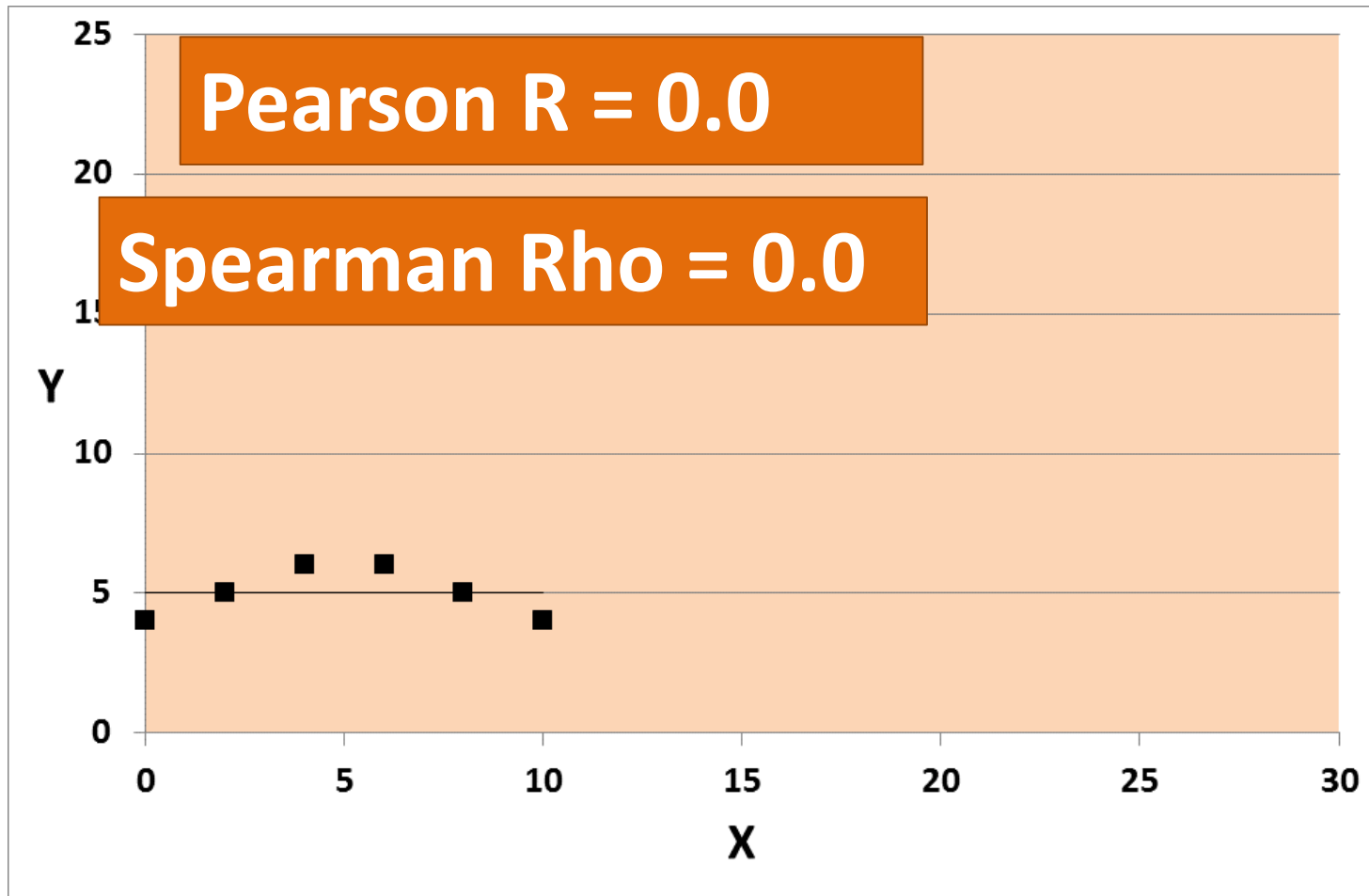
- **TR/JCR rankings of journals by impact factor and subject category are widely available**
- **Differences in means and positive correlations are often misunderstood**



## Study by Sugimoto et al (2014)

- The rejection rate of manuscripts submitted to a journal shows a **significantly positive** correlation with the journal's impact factor
- Journals with rigorous referee systems tend to generate higher impact than others
- Then they divided journals into **quartiles** according to their rejection rates
- **Within quartiles**, the correlation coefficient was **lower and not significant**

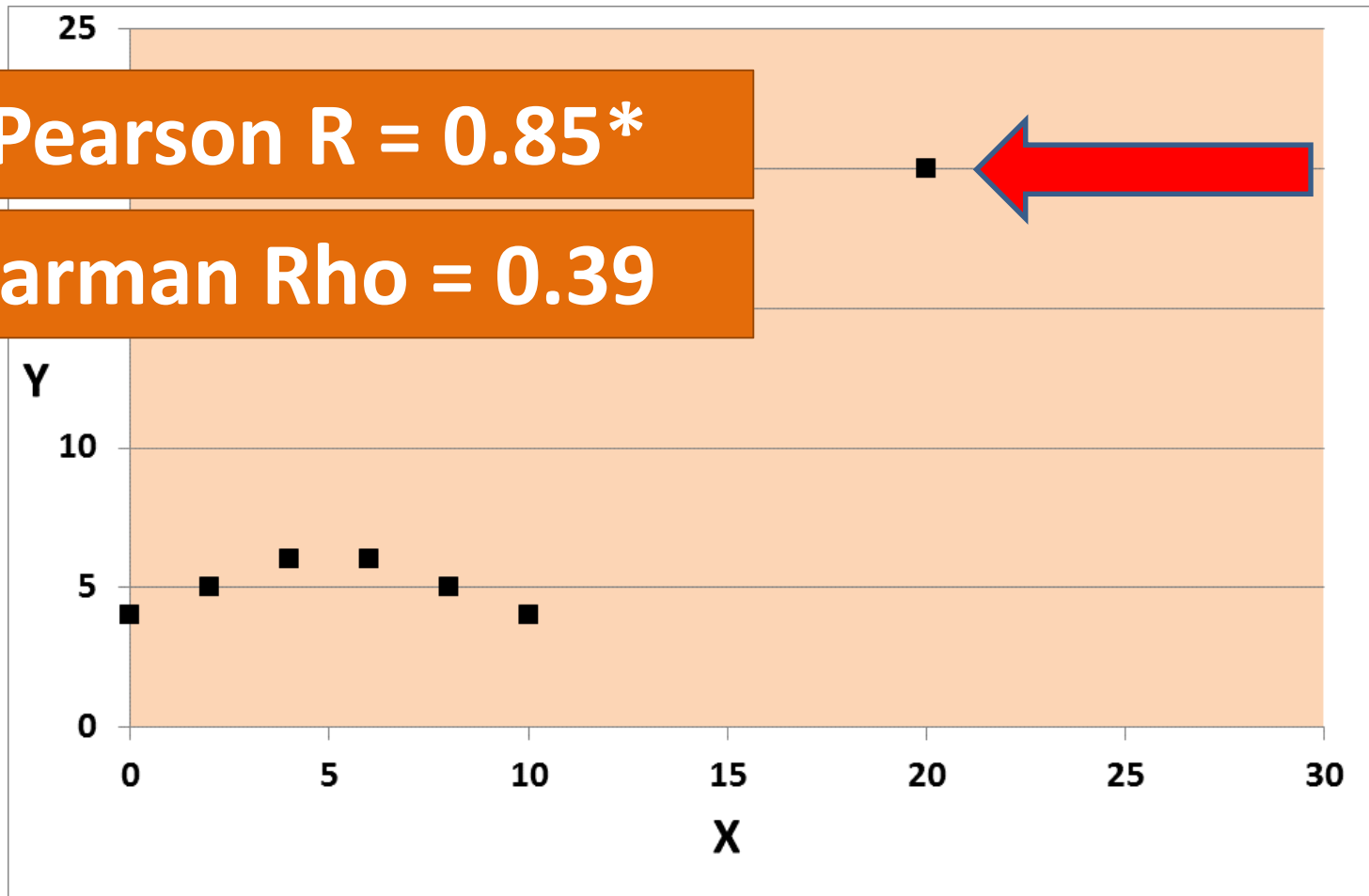
# Linear / rank correlations - Case 1



# Linear / rank correlations - Case 2

Pearson R = 0.85\*

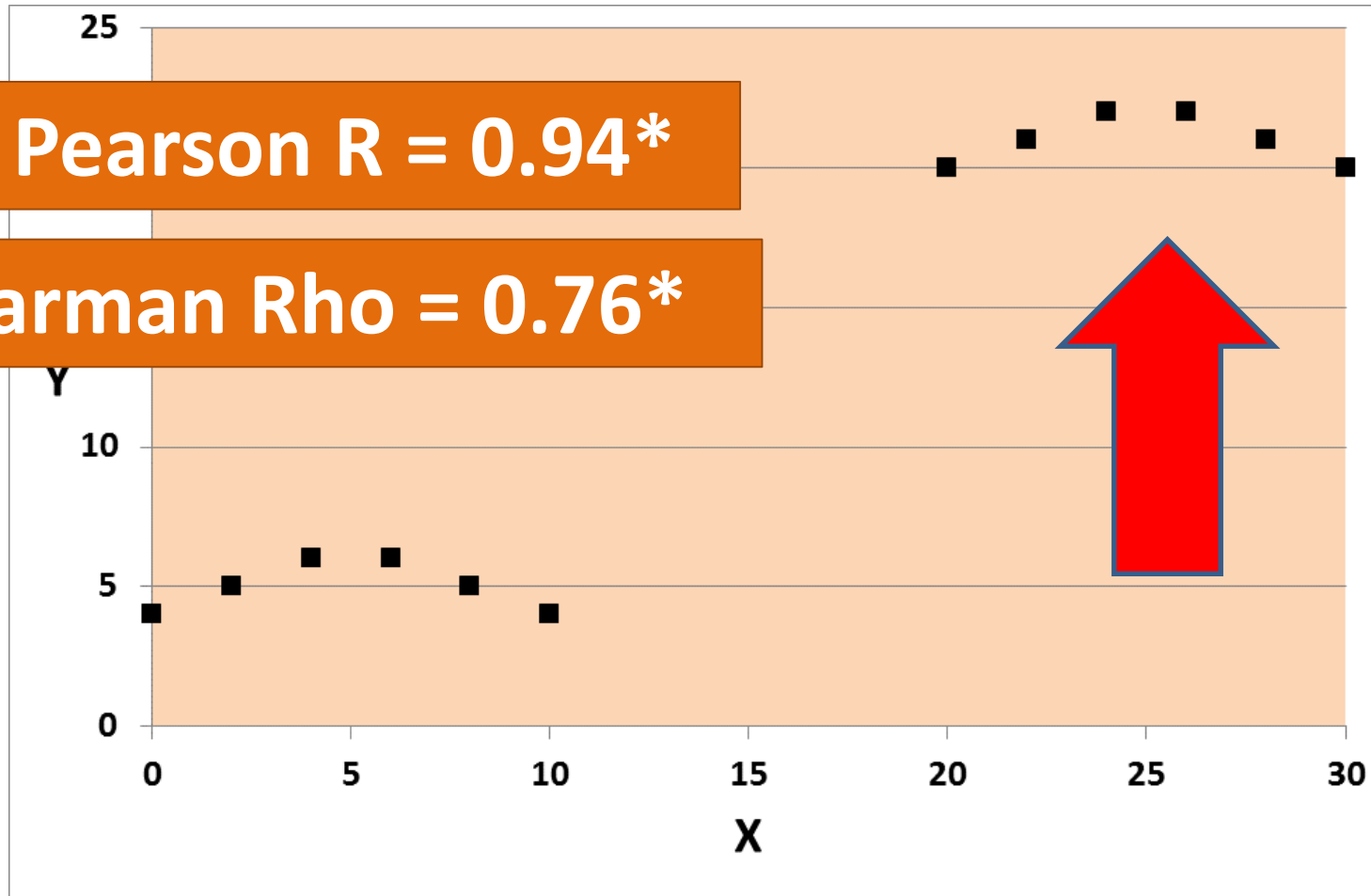
Spearman Rho = 0.39



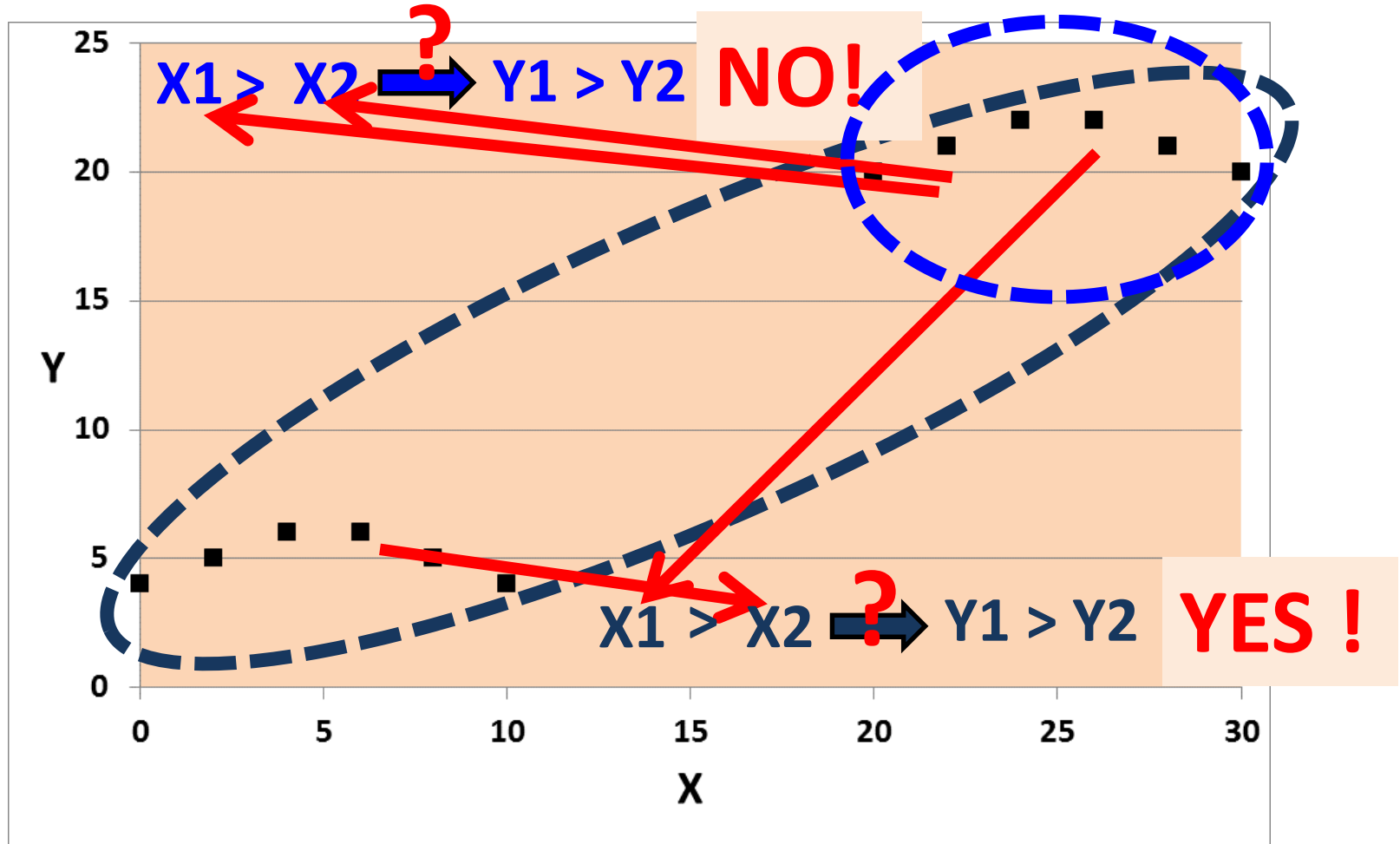
# Linear / rank correlations - Case 3

Pearson R = 0.94\*

Spearman Rho = 0.76\*



# Linear / rank correlations - Case 3



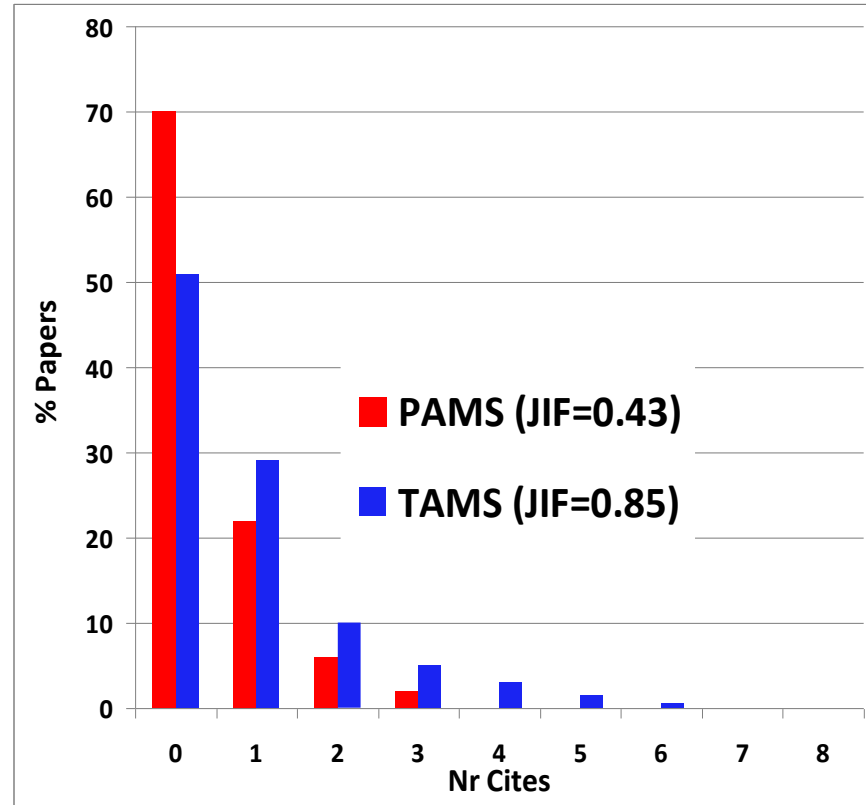
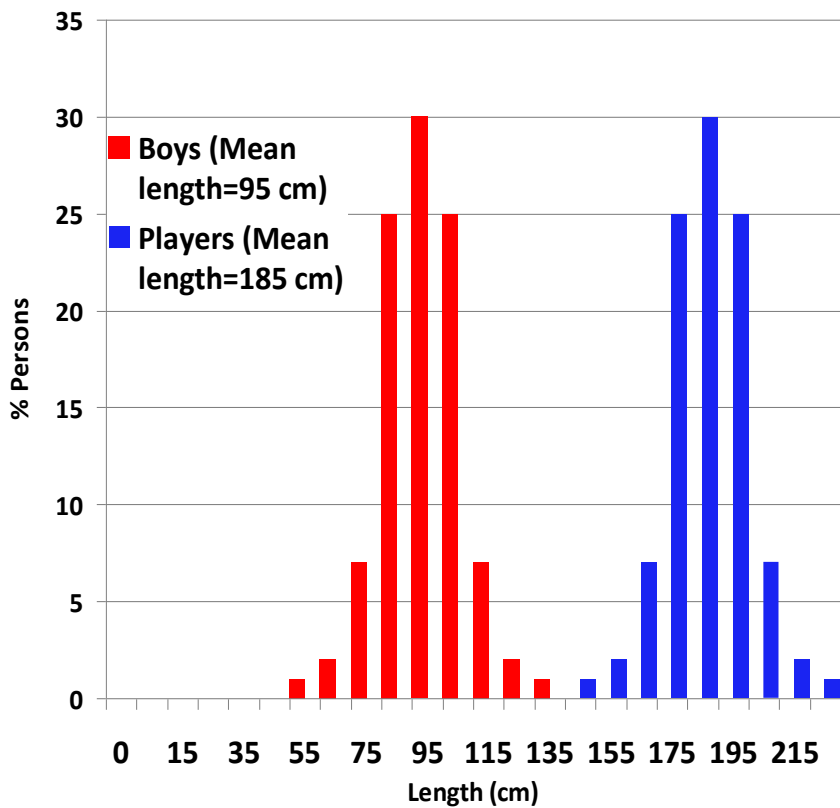
**ISI/JCR Journal Impact Factor  
of journal J for year T**

**Citations in year T to items published in J in  
years T-1 and T-2**

**÷**

**Number of “citable” items published in J in  
years T-1 and T-2**

# Normal vs. skewed distributions



# What is the probability that .....

**a randomly selected boy  
is at least as tall as a  
randomly selected adult?**

**Av. Length: Boys 85 cm; Adults: 185 cm**

**Almost zero**

**a randomly selected PAMS paper  
is cited at least as often as a randomly  
selected TAMS paper?**

**JIF: PAMS: 0.43; TAMS: 0.85**

**62 %**



# How to use Scopus as tool in research assessment

- Scopus is **not** constructed according to Eugene Garfield's principles
- Scopus covers 'national' journals, many from **scientifically developing** countries
- **25 per cent of journals** in Scopus receives **at most 15 citations** (in RIP impact factor window) and **10 per cent** at most 3 citations (data for 2011).
- 25 per cent of Scopus journals has an impact factor (RIP) **below 0.21**.
- **Consider:** Disregard the bottom 10 % or even bottom quartile of journals

# How to use journal impact measures

1. Do **not** use “**raw**” journal impact factors
2. Use **field-normalized** indicators ( e.g., “relative journal impact indicator”, SNIP, percentile ranks)
3. Do **not** use the number of publications in **absolute top journals (e.g., top 3)** as an indicator
4. Rather focus on:
  - a) **Number of articles in top quartile** (“top 25 per cent”) in a subject field (or groups of subject fields containing at least 50 journals), or even in top 50%.
  - b) **Diversity, e.g., number of different journals**

## **5. Role of journal impact factors must be reduced**

- The DORA Declaration makes sense: their role in promoting journals must be reduced**
- The genuine alternative is use of metrics of the quality of journals' peer review system**

## **6. More advanced indicators are technically feasible**

- **A challenge: The practical realization of the ethos of science and scholarship in a digital age**
- **Development of indicators has not yet fully profited from the computerization of the research process.**

# Computerization of the research process

**Information processing**

**Searching, browsing, reading behavior**

**Usage data (e.g. downloads) patterns**

**Communication & organization**

**Online tools amplifying collective intelligence**

**Use of blogs, wikis, issue trackers**

**Research assessment**

**Scholars' interaction with research outputs**

**Mentions in social media & ref managers**

**Research data & methodologies**

**Big data; crowdsourcing; analysis software**

**Analysis top cited articles by subject field**

# Comp Sci-related top-cited articles in Scopus [G. Halevi in Res Trends]


# Cites	Discipline	Article Title
17,171	Agr & Biol Sci, Mol Biol; Medicine	<b>MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 (2007)</b>
4,335	Business, managemt, social sciences	<b>User acceptance of information technology: Toward a unified view (2003)</b>
5,325	Chemistry	<b>UCSF Chimera - A visualization system for exploratory research and analysis (2004)</b>
15,191	Computer Sci; Eng	<b>Distinctive image features from scale-invariant keypoints (2004)</b>
1,335	Energy	<b>Geant4 developments and applications (2006)</b> [software for simulating passage of particles through matter]
7,784	Engineering; Math	<b>A fast and elitist multi-objective genetic algorithm: NSGA-II (2002)</b>
4,026	Environm Sci	<b>GENALEX 6: Genetic analysis in Excel. Population genetic software....(2006)</b>
4,404	Materials Science	<b>The SIESTA method for ab initio order-N materials simulation (2002)</b>
10,921	Physics & Astron	<b>Coot: Model-building tools for molecular graphics (2004)</b>

## **7. Advanced online self-assessment tools are needed**

- An advanced online tool for bibliometric self-assessment is technically feasible....**
- ....combining Eugene Garfield's algorithm and Robert K. Merton's notion of a reference group**

# Researcher Bibliometric Self-Assessment Tool

1. Select, **verify**, add, correct, finalize the **list of publications** of the assessed researcher (AR) (SET 1)



2. Select a set (SET 2) of articles that are **similar** to those in SET1 in terms of **subject field**




3. Identify **the authors** in SET2 and **all** their articles (SET 3)



4. Select the authors in SET3 most similar to the AR in terms of **subject field** and **academic age** (SET 4)



5. Let the user further select the authors in SET 4 he wants to include in his tailormade benchmark set (SET 5)



6. Calculate **quartile**-based indicators for the AR using the author data in the author benchmark SET 5



# Output of the Researcher Bibliometric Self-Assessment Tool

## Step 6:

Author assessment: Comparing Assessed Author with Related (=benchmark) Authors

Assessed Author: | xxx

Indicators are based on the network benchmarking

Indicator	Assessed Author		Related/benchmark authors				
	SCORE	QUARTILE (1=Bottom)	N	MEAN	Q1	MEDIAN	Q3
Articles	65	3	124	109.4	22.0	41.5	82.0
Citations	192	2	124	520.2	79.5	215.0	482.5
Cites/Article	3.0	1	124	5.2	3.1	4.7	6.9

3-level indicators

The evaluated author scores in the 1st quartile in the number of articles, 2nd in the number of citations and 3rd in the cites per article compared to its peers

## 8. Manuscript peer review is essential

- Distinguish two dimensions:
  1. Applying base-line quality criteria – facilitating manuscript improvement
  2. Focusing on groundbreaking findings – facilitating scientific debate

## **9. Focus not merely on output, also on input and process**

- **Efficiency measures relate output to input**
- **Science of Team Science (SCITS) studies conditions for successful collaboration**

## **10. Metrics should be independent and have wide access**

- Metrics development must apply rigorous scientific quality standards independent of data providers, funders and politicians**
- Metrics availability should not be controlled by business or political interests**

**Thank you for your attention**