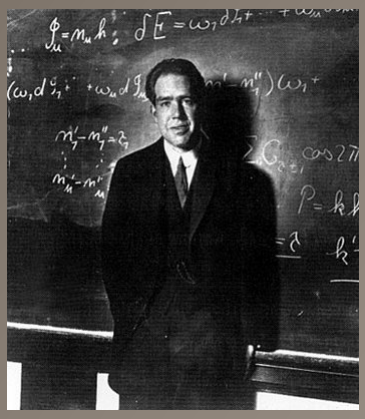
THE ART & SCIENCE OF IDENTIFYING FUTURE NOBEL LAUREATES

Discover the Power of Quantitative Analysis
The Art & Science of Identifying Future Nobel Laureates

DAVID PENDLEBURY
Consultant
Healthcare & Science
November 2009

THOMSON REUTERS



"Prediction is very difficult, especially if it's about the future."

-- Niels Bohr, Nobel Laureate in Physics, 1922

THOMSON REUTERS

Successful predictions since 2002

- 2002: Sydney Brenner, Medicine (predicted in 1989); Daniel Kahneman, Economics (predicted in 2002)
- 2003: Robert F. Engle, Economics (predicted in 2003) and Clive W.J. Granger, Economics (predicted in 2003)
- 2004: David J. Gross, Physics (predicted in 1990) and Frank A. Wilczek, Physics (predicted in 1990)
- 2005: Robert H. Grubbs, Chemistry (predicted in 2005)
- 2006: Edmund S. Phelps, Economics (predicted in 1990)
- 2007: Mario Capecchi, Medicine (predicted in 2006) and Martin J. Evans, Medicine (predicted in 2006) and Oliver Smithies, Medicine (predicted in 2006); Albert Fert, Physics (predicted in 2006) and Peter Grünberg, Physics (predicted in 2006)
- 2008: Paul Krugman, Economics (predicted in 2006); Roger Y. Tsien, Chemistry (predicted in 2008); Luc Montagnier, Medicine (predicted in 1989)
- 2009: Elizabeth H. Blackburn, Medicine (predicted in 2009) and Carol W. Greider, Medicine (predicted in 2009), and Jack W. Szostak, Medicine (predicted in 2009); Oliver E. Williamson, Economics (predicted in 2006)

THOMSON REUTERS

Odds against forecasting future Nobel Prize winners

Nobel Prize in…	Est. Number of researchers		
	100%	1%	.1%
Medicine	2,000,000	20,000	2,000
Chemistry	1,000,000	10,000	1,000
Physics	1,000,000	10,000	1,000
Economics	200,000	2,000	200

Total number of Nobel Laureates since 1901 => Medicine 195; Chemistry 156; and Physics 186; Economics (since 1969) 64

THOMSON REUTERS

What accounts for this success?

Psychic powers?

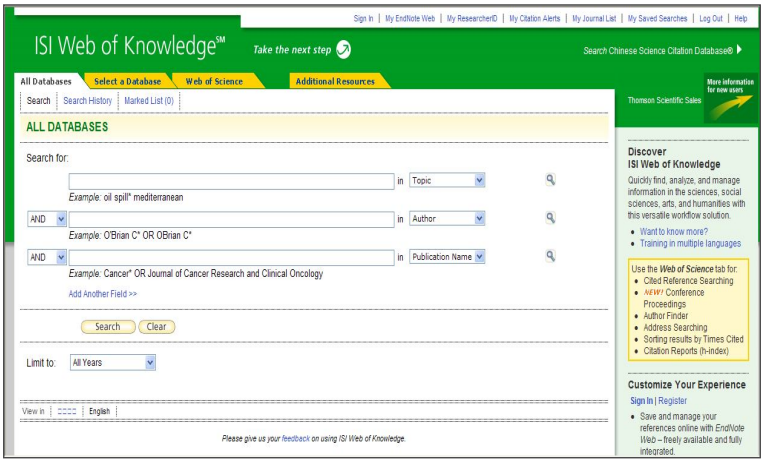
THOMSON REUTERS

And certainly not me!

No expert or special knowledge

THOMSON REUTERS

This is the answer…



THOMSON REUTERS

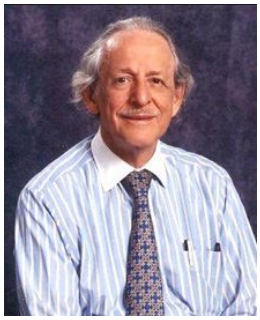
ISI Web of Knowledge, including the Web of Science

- Rigorous selection of internationally influential journals, continually reviewed
- Comprehensive in respect to complete information on each record – with citations
- Unmatched time depth, 1900 to present: 44MM records and 700MM citations
- Powerful retrieval and analytic capabilities




THOMSON REUTERS

…and the person responsible for it



Eugene Garfield

Pioneer Information Scientist

and Inventor of Citation Indexes for
Sciences, Social Sciences & Humanities

THOMSON REUTERS

Citation Indexing: Retrieval and Analysis

- Eugene Garfield, "Citation indexes for science: a new dimension in documentation through association of ideas," Science 122(3159):108-11, 15 July 1955
- New method of information retrieval for sciences, although used in legal literature (Shepard's Citations)
- Avoids lexical problems in indexing and retrieval
- Allows searching "forward in time"
- Reveals socio-cognitive connections of researchers, as well as the organic structure and hierarchy of science
- Permits quantitative investigations and descriptions of research enterprise – "turning the tools of science on science itself," Derek de Solla Price (1963)

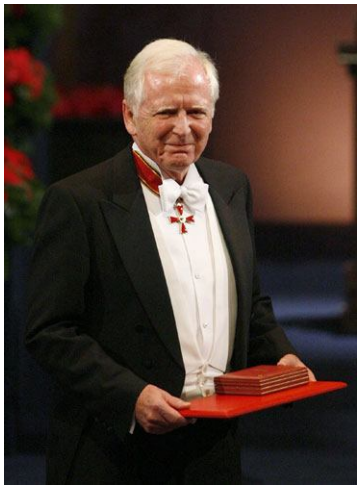
THOMSON REUTERS

Outline

- The Nobel Prize: Selection process
- Citation data and correlation between high frequency and peer esteem
- Thomson Reuters efforts and methods
- Meaning for science administrators and policymakers

THOMSON REUTERS

The Nobel Prize, 1901- : Selection Process



- Alfred Nobel's will: 1895
- Prizes in Physiology or Medicine, Chemistry, Physics, Literature, and Peace (Economics Prize was begun in 1969)
- Nomination process, deliberations by committees, selection
- Limit of three winners for each Nobel Prize

THOMSON REUTERS

Eugene Garfield, "Citation Indexing for Studying Science," ***Nature***, 227, pp. 669-671, August 15, 1970

50 Most Cited Authors for 1967, primary author citations

1	LOWRY OH	2921	26	ELIEL EL	721
2	CHANCE B	1374	27	STREITWIESER A	717
3	**LANDAU LD (1962)**	1174	28	**MULLIKEN RS (1965)**	712
4	**BROWN HC (1979)**	1150	29	**JACOB F (1965)**	711
5	**PAULING L (1954)**	1063	30	**BORN M (1954)**	710
6	**GELL-MANN M (1969)**	942	31	BRACHET J	706
7	COTTON FA	940	32	WINSTEIN S	702
8	**POPLE JA (1998)**	933	33	ALBERT A	687
9	BELLAMY LJ	906	34	LUFT JH	674
10	SNEDECOR GW	904	35	**DE DUVE C (1974)**	673
11	**BOYER PD (1997)**	893	36	**VON EULER US (1970)**	668
12	BAKER BR	876	37	FIESER LF	666
13	KOLTHOFF IM	853	38	HUISGEN R	667
14	**HERZBERG G (1971)**	842	39	NOVIKOFF AB	655
15	FISCHER F	826	40	GOODWIN TW	643
16	SEITZ F	822	41	**BARTON DHR (1969)**	632
17	DJERASSI C	801	42	FISHER RA	631
18	BERGMEYER HU	754	43	BATES DR	627
19	WEBER G	750	44	**FLORY PJ (1974)**	626
20	REYNOLDS ES	748	45	STAHL E	626
21	**MOTT NF (1977)**	741	46	DEWAR MJS	619
22	**ECCLES JC (1963)**	737	47	GILMAN H	618
23	FEIGL F	729	48	FOLCH J	618
24	FREUD S	727	49	DISCHE Z	614
25	PEARSE AGE	726	50	GLICK D	609

THOMSON REUTERS

Eugene Garfield, "Who Will Win the Nobel Prize in Economics? Here's a Forecast Based on Citation Indicators," *Current Comments*, No. 11, pp. 3-7, March 12, 1990

Most Cited Economists, *SSCI* 66-86, ranked by primary author citations

1	**ARROW KJ**	**1972**	26	HOTELLING H	d.1973
2	**SAMUELSON PA**	**1970**	27	MINCER J	d. 2006
3	**SIMON HA**	**1978**	28	**COASE RH**	**1991**
4	**FRIEDMAN M**	**1976**	29	NERLOVE M	
5	**BECKER GS**	**1992**	30	**DEBREU G**	**1983**
6	FAMA EF ←2002		31	JORGENSON DW ←2006	
7	FELDSTEIN M ←2008		32	ZELLNER A	
8	THEIL H	d. 2000	33	**SCHULTZ TW**	**1979**
9	**STIGLER GJ**	**1982**	34	**PHELPS ES**	**2006**
10	BAUMOL WJ		35	BLACK F	d. 1995
11	**BUCHANAN JM**	**1986**	36	**STIGLITZ JE**	**2001**
12	GALBRAITH JK	d. 2006	37	OLSON M	d. 1998
13	**TOBIN J**	**1981**	38	**KLEIN LR**	**1980**
14	KEYNES JM	d. 1946	39	MALINVAUD E	
15	**MODIGLIANI F**	**1985**	40	LINTNER J	d. 1984
16	BARRO RJ ←2002		41	**GRANGER CWJ**	**2003**
17	ROBINSON J	d. 1983	42	JENSEN MC	
18	**HICKS JR**	**1972**	43	MUSGRAVE RA	d. 2007
19	**LUCAS RE**	**1995**	44	BHAGWATI JN ←2006	
20	**SEN AK**	**1998**	45	ALCHIAN AA ←2008	
21	**MYRDAL G**	**1974**	46	MANSFIELD E	d. 1997
22	**SOLOW RM**	**1987**	47	**KUZNETS S**	**1971**
23	GRILICHES Z	d.1999	48	CHOW GC	
24	SARGENT TJ ←2008		49	HIRSHLEIFER J	d. 2005
25	BOWLES S		50	CHENERY HB	d. 1994

Select Essays on Forecasting Nobel Prize Winners using Citation Data

1965

Sher I.H. and Garfield E, "New tools for improving and evaluating the effectiveness of research," in M.C.Yovits, D.M. Gilford, R.H. Wilcox, E. Staveley and H.D. Lemer, eds., Research Program Effectiveness, Proceedings of the Conference Sponsored by the Office of Naval Research Washington, D.C., July 27-29, 1965 (New York: Gordon and Breach, 1966) p.135-146.

1968

Garfield E. and Malin M.V. "Can Nobel Prize winners be predicted? " p.1-17, 1968. Presentation (Unpublished), No:166. Paper presented at 135th Annual Meeting, AAAS, Dallas, TX, USA. December 26-31, 1968

1970

Garfield E., "Citation indexing for studying science," Nature, 227, pp. 669-671, August 15, 1970

1992

Garfield E. and Welljams-Dorof A., "Of Nobel Class: a citation perspective on high impact research authors," Theoretical Medicine, 13(2): 117-35, 1992. *** Essential

2006

Garfield E., "Identifying Nobel Class scientists and the uncertainties thereof," Presentation at the European Conference on Scientific Publication In Medicine and Biomedicine, Lund, Sweden, April 21, 2006, and at the Third Nordic Conference on Scholarly Communication, Lund, Sweden, April 25, 2006.

GARFIELD, E and WELLJAMS-DOROF, A. "Of Nobel Class - a citation perspective on high-impact research," THEORETICAL MEDICINE, 13 (2): 117-135, June 1992

The purpose of this paper was to determine if quantitative rankings of highly cited research authors confirm Nobel prize awards. Six studies covering different time periods and author sample sizes were reviewed. The number of Nobel laureates at the time each study was published was tabulated, as was the number of high impact authors who later became laureates. The Nobelists and laureates-to-be were also compared with non-Nobelists to see if they differed in terms of impact and productivity. The results indicate that high rankings by citation frequency identify researchers of Nobel class - that is, a small set of authors that includes a high proportion of actual Nobelists and laureates-to-be. Also, the average impact (citations per author) of Nobelists and laureates-to-be is sufficiently high to distinguish them from non-Nobelists in these rankings. In conclusion, a simple, quantitative, and objective algorithm based on citation data can effectively corroborate - and even forecast - a complex, qualitative, and subjective selection process based on human judgment.

Who is “of Nobel Class”?

- **Harriet A. Zuckerman**, Scientific Elite: Nobel Laureates in the United States, 1977 (also 1996)
- “Every year more scientists are eligible for the Nobel Prizes than can win them….There has always been an accumulation of uncrowned laureates who are peers of the prize-winners in every sense except that of having the award
- The French Academy which has 40 members – the Immortals. “Who shall occupy the 41st chair?”

THOMSON REUTERS

Another example: Chemistry, 1981-June 1997

Simple list of top 50 chemists by citations included seven Nobelists and six more who later received the Nobel Prize in Chemistry

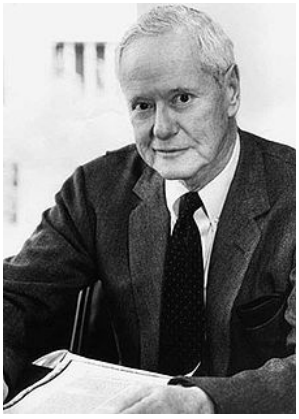
Data included 627,871 unique author names; therefore, top 50 are in top .01%

Rank	Name	Year	Rank	Name	Year
4	RR Ernst	1991	2	JA Pople	1998
10	JM Lehn	1987	33	AH Zewail	1999
12	RE Smalley	1996	37	KB Sharpless	2001
16	EJ Corey	1990	40	RR Schrock	2005
32	R Hoffmann	1981	46	G Ertl	2007
41	HW Kroto	1996	48	R Noyori	2001
50	DHR Barton	1969			

THOMSON REUTERS

Theories of Citation

- Robert K. Merton, (1910-2003), sociologist of science, Columbia University. Normative theory.
- Citations as currency used to repay intellectual debts. Those with many citations have gained "credits" from their peers.
- The formal nature of publication and the moral imperative to cite.
- Other theories, including citations as rhetorical devices, constructivist theories, "schools"



THOMSON REUTERS

Validation studies: Citations frequency and its correlation with other measures of peer esteem

Name	Field	Year
Kenneth E. Clark	Psychology	1957
Jonathan R. Cole & Stephen Cole	Physics	1967, 1973
Henry G. Small	Collagen research	1977
Julie A. Virgo	Cancer research	1977
Michael E.D. Koenig	Pharmaceutical research	1983
Eugene Garfield	Nobel Prize winners	1992
Charles Oppenheim	University rankings (RAE)	Mid 1990s-
Andy T. Smith & Michael Eysenck	Psychology	2002

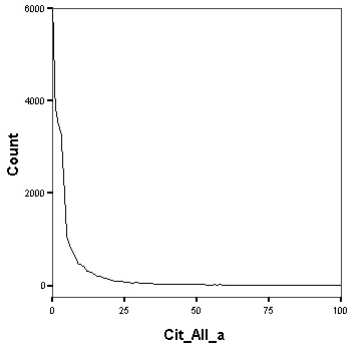
Typical findings, r = .7 to .9

Smith and Eysenck, comparing 1996 and 2001 RAE scores given to psychologists at 38 UK universities (peer review) with their citation counts, concluded:

"The two approaches measure broadly the same thing."

THOMSON REUTERS

Skewed distributions persistent at all levels, ubiquitous in information science



- Lotka, Bradford, Zipf and numerous variant "laws"
- Pareto: Income inequality.

  The "80:20" Rule
- Author productivity, citations; also true for papers, journals, institutions, nations, etc.
- Merit: "Some things are better than others"
- Human behavior, judgment, and selection. Price and Merton's cumulative advantage
- Skewed distribution is an aid in identifying those of Nobel Class

THOMSON REUTERS

Qualitative vs. Quantitative: Two (complementary) types of peer review

Peer Review: Qualitative

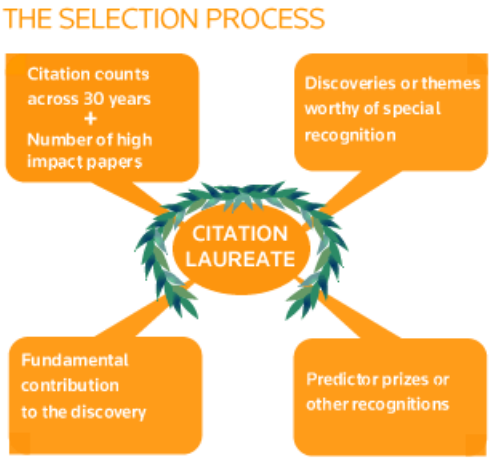
- Small-scale, ground-up view
- Absolute counts, size colors perceptions, judgments
- Affected by work done long ago

Citation Analysis: Quantitative

- Global, top-down view
- Weighted and relative measures
- Can reveal more recent contributions

THOMSON REUTERS

Thomson Reuters procedure



THOMSON REUTERS

1. Choosing types of citation analysis

- Total citations, generally over three or more decades, and within field, since different fields exhibit different average rates of citation. Differences as much as 10 times (mathematics vs. molecular biology)
- Citations per paper, or citations per author, within field
- Multiple highly cited papers (the h-index as alternative)
- Authorship of highly cited review articles
- Authorship of "core" or foundation papers in research fronts, constructed through co-citation analysis
- Time distribution of citations – looking for a rising pattern
- Citations focused on a discovery rather than for lifetime of work on many different areas of research in a field

THOMSON REUTERS

2. Assessing the research

- Looking beneath the citation counts: Is the citation record focused or dispersed among many contributions?
- Highly cited papers: One or many? Methods papers, review papers, co-authorship but not primary contributor, as well as papers with hundreds of authors?
- What is the discovery? What is its importance in the field? Is it a fundamental advance or a revolutionary finding? Or does it extend fundamental work?
- Has the area of research already been recognized with a Nobel Prize, especially recently?
- How does the contribution compare to others deemed prize-worthy?

3. Discerning the key contributors

- The limit, per Nobel's will, of not more than three winners for each Prize
- The Nobel committee's preference for selecting pioneering discoveries rather than those who extended the work, however importantly
- Questions of priority. Questions of controversy. The discovery of HIV in 1983-1984 as an example: Robert Gallo vs. Luc Montagnier
- Nobel Prizes rarely reveal errors of commission but always, and out of necessity, are subject to errors of omission. Conservative choices

4. Complementary indicators: Other evidence of peer esteem

- Prizes, especially "predictor prizes," such as the Lasker, Gairdner, Wolf, Japan; also Kyoto, Harvey, Crafoord, Shaw, Franklin, Gruber, and Bates (for economics); 5X increase in 20 years
- National academy memberships, foreign academy memberships, special societies
- Keynote speeches at national and international conferences
- Senior positions at universities, head large labs
- Editorship of high-impact journals

THOMSON REUTERS

Naming Citation Laureates

- The pool of potential Nobel Prize winners, occupants of the 41st chair:

- "Of Nobel Class"…. like Nobel Laureates in even way except having won the Prize.
- Since Thomson Reuters selects them using citation measures primarily, we name them Citation Laureates

THOMSON REUTERS

Medicine Citation Laureates, 2009

- **Elizabeth H. Blackburn**, Morris Herztein Professor of Biology and Physiology, Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, USA, and
- **Carol W. Greider**, Daniel Nathans Professor and Director, Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore, MD, USA, and
- **Jack W. Szostak**, Professor of Genetics, Harvard Medical School, and Alexander Rich Distinguished Investigator at Massachusetts General Hospital, Boston, MA, USA; also, Howard Hughes Medical Institute Investigator

"for their roles in the discovery of and pioneering research on telomeres and telomerase"

Medicine Citation Laureates, 2009

- **James E. Rothman**, Fergus F. Wallace Professor of Biomedical Sciences, Professor and Chairman of Cell Biology, Professor of Chemistry, Yale University, New Haven, CT, USA, and
- **Randy Schekman**, Professor of Cell and Developmental Biology, University of California Berkeley, Berkeley, CA, USA; also, Howard Hughes Medical Institute Investigator

"for their research on cellular membrane trafficking"

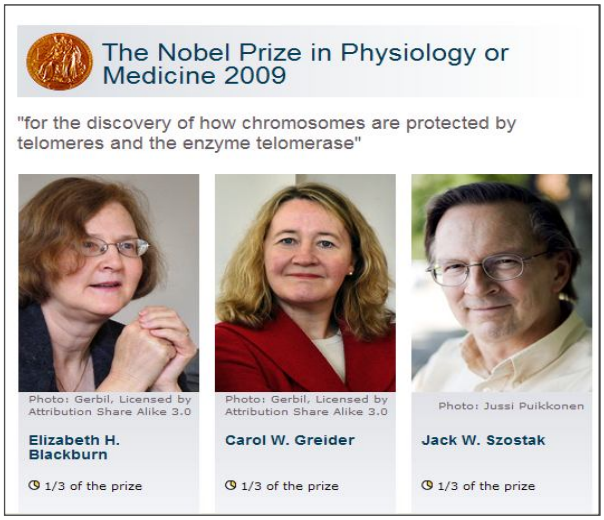
Medicine Citation Laureate, 2009

- **Seiji Ogawa**, Director, Ogawa Laboratories for Brain Function Research, Hamano Life Science Research Foundation, Tokyo, Japan

"for his fundamental discoveries leading to functional magnetic resonance imaging (fMRI), which has revolutionized basic research in brain science and diagnosis in clinical medicine"

THOMSON REUTERS

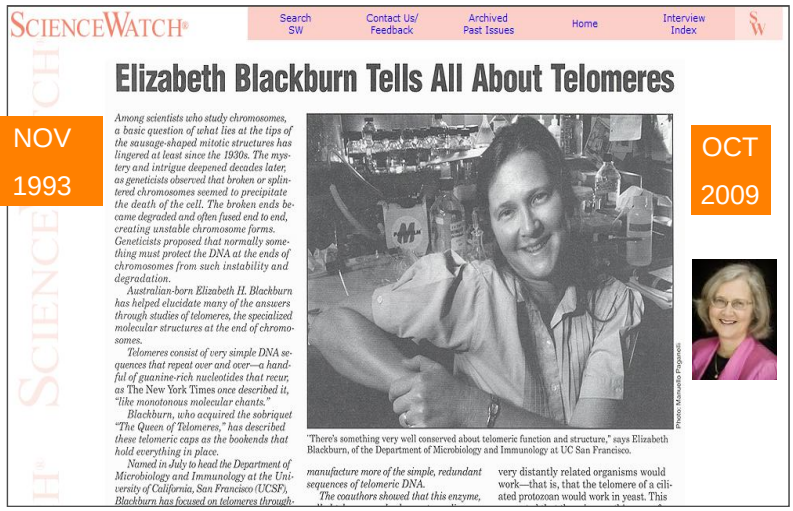
And the winners…



THOMSON REUTERS

Tracking Blackburn's citations since early 1990s

NOV 1993



SCIENCEWATCH®

Search SW | Contact Us/ Feedback | Archived Past Issues | Home | Interview Index

Elizabeth Blackburn Tells All About Telomeres

Among scientists who study chromosomes, a basic question of what lies at the tips of the sausage-shaped mitotic structures has lingered at least since the 1930s. The mystery and intrigue deepened decades later, as geneticists observed that broken or splintered chromosomes seemed to precipitate the death of the cell. The broken ends became degraded and often fused end to end, creating unstable chromosome forms. Geneticists proposed that normally something must protect the DNA at the ends of chromosomes from such instability and degradation.

Australian-born Elizabeth H. Blackburn has helped elucidate many of the answers through studies of telomeres, the specialized molecular structures at the end of chromosomes.

Telomeres consist of very simple DNA sequences that repeat over and over—a handful of guanine-rich nucleotides that recur, as The New York Times *once described it, "like monotonous molecular chants."*

Blackburn, who acquired the sobriquet "The Queen of Telomeres," has described these telomeric caps as the bookends that hold everything in place.

Named in July to head the Department of Microbiology and Immunology at the University of California, San Francisco (UCSF), Blackburn has focused on telomeres through-

"There's something very well conserved about telomeric function and structure," says Elizabeth Blackburn, of the Department of Microbiology and Immunology at UC San Francisco.

manufacture more of the simple, redundant sequences of telomeric DNA.

The coauthors showed that this enzyme,

very distantly related organisms would work—that is, that the telomere of a ciliated protozoan would work in yeast. This

Photo: Manuello Pogiandi

OCT 2009

THOMSON REUTERS

Citations: Lagging indicators as to research, but leading indicators as to peer esteem, prizes

Interviews in *Science Watch*	**Nobel Prize**
Franck Wilczek, April 1991	Physics, 2004
Ahmed Zewail, July 1991	Chemistry, 1999
Harold Kroto, January 1992	Chemistry, 1996
F. Sherwood Rowland, May 1992	Chemistry, 1995
Mario Capecchi, July/August 1993	Medicine, 2007
Elizabeth H. Blackburn, November 1993	Medicine, 2009
K. Barry Sharpless, October 1995	Chemistry, 2001
Wolfgang Ketterle, January/February 1999	Physics, 2001
	Ave. 8.25 years

THOMSON REUTERS

Are Nobel Prizes good indicators of national performance in science?

Christopher T. Hill, The Nobel-prize Awards in Science as a Measure of National Strength in Science (U.S. Congressional Research Service, 1986)

Tibor Braun, Zsuzsa Szabadi-Peresztegi, and Eva Kovacs-Nemeth, "No-bells for ambiguous lists of ranked Nobelists as science indicators of national merit in physics, chemistry and medicine, 1901-2001," Scientometrics, 56 (1): 3-28, 2003

- A VERY lagging indicator – sometimes of several decades: "Counting the number of Nobel Prizes…may be only recording extinct volcanoes" – Jacques Barzun
- A "thin" indicator. Few prizes. Nationality questions. Measurement questions: per capita? Also, the peculiarities of the Nobel Prize nomination process
- Better assessments of national research performance should use more recent, broader indicators.

THOMSON REUTERS

"The goal in science…"

"The goal in science should be to find out things – not to win a prize….Yet, in conversations with some of my younger colleagues, I get a sense that it has become a goal, and that is not good. I think it would be better if there were no prizes."

-- Salvador Luria, Nobel Laureate in Physiology or Medicine, 1969

THOMSON REUTERS

Nobel winners, Everyone else

THOMSON REUTERS

Citation analysis and research evaluation: National to individual

Henk F. Moed, Citation Analysis in Research Evaluation,

(Springer 2005)

General Principles

- Basic vs. applied science
- Aggregate vs. individual
- Long time vs. short time
- Relative vs. absolute measures
- Multiple measures vs. single measures

Above all: Compare like with like,

not “apples with oranges”

THOMSON REUTERS

Specific observations

- Demand for single measures in evaluation is unreasonable, antithetical to good evaluation (h-index)
- Use of journal impact factors as surrogate measures of individual achievement is misleading (use actual citations to papers, researchers, institutions, nations)
- Citation analysis is not “a royal road to evaluation.” It is not a short-cut method, but actually requires more work, more thought
- Quantitative results should always be combined with qualitative judgments (peer review). Each reinforces the other. Nobel forecasts benefit from the combination of quantitative data with other measures of peer esteem (such as prizes)

For policymakers and administrators: Using data for retrospective analysis vs. for future rewards

- Law of Unintended Consequences, resulting in a negative effect contrary to what was intended. Can stem from perverse incentives, short- over long-term goals
- Goodhart’s Law (1975): “Once a social or economic indicator or other surrogate measure is made a target for the purpose of conducting social or economic policy and control, then it will lose the information content that would qualify it to play such a role.”
- In both cases, setting a simple or crude measure of performance changes behavior as subjects attempt to optimize their performance – not only does this disturb behavior, it also destroys the utility of the measure
- Goal in science is not citations and not prizes: the goal is excellence in research. Citations and prizes will follow

Scientists, Policymakers, Citizens

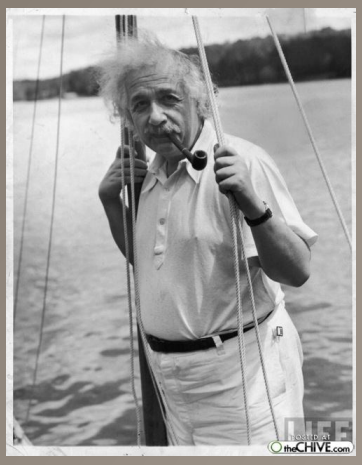
- Support for science from taxpayers requires policymakers and administrators to ensure effectiveness and efficiency. Scientists are accountable for support.
- Citation analysis combined with peer review can add substantially to research assessment and improve decisions made by administrators, policymakers
- But using metrics in simple ways to control outcomes can change behavior and actually institutionalize uniformity or even mediocrity in research
- This may certainly dampen creativity and derail “revolutionary science” (Kuhn), the type that is recognized as excellent by the Nobel Prize committees
- The ideal is informed, thoughtful, and wise assessment coupled with directed (but not overly directed) support related to national or institutional goals.

THOMSON REUTERS

A little wisdom…

"Not everything that counts can be counted, and not everything that can be counted counts."

From a sign hanging in Albert Einstein's office at the Institute for Advanced Study in Princeton.



THOMSON REUTERS

Citation Analysis: Select Bibliography

- **Eugene Garfield**, Citation Indexing—Its Theory and Application in Science, Technology and Humanities, John Wiley & Sons, 1979
- **Henk F. Moed**, Citation Analysis in Research Evaluation, Springer, 2005
- **Nicola De Bellis**, Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics, Scarecrow Press, 2009

Contact Information for David Pendlebury

Email: david.pendlebury@thomsonreuters.com

Goodbye from Oregon.
Thank you.

THOMSON REUTERS