Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 72, No. 3 (2007) 513–520
DOI: 10.1007/s11192-007-1573-7

The difference between highly and poorly cited medical articles in the journal *Lancet*

RONALD N. KOSTOFF

Office of Naval Research, Arlington, VA (USA)

Characteristics of highly and poorly cited research articles (with Abstracts) published in *The Lancet* over a three-year period were examined. These characteristics included numerical (numbers of authors, references, citations, Abstract words, journal pages), organizational (first author country, institution type, institution name), and medical (medical condition, study approach, study type, sample size, study outcome). Compared to the least cited articles, the most cited have three to five times the median number of authors per article, fifty to six hundred percent greater median number of references per article, 110 to 490 times the median number of citations per article, 2.5 to almost seven times the median number of Abstract words per article, and 2.5 to 3.5 times the median number of pages per article.

The most cited articles' medical themes emphasize breast cancer, diabetes, coronary circulation, and HIV immune system problems, focusing on large-scale clinical trials of drugs. The least cited articles' themes essentially do not address the above medical issues, especially from a clinical trials perspective, cover a much broader range of topics, and have much more emphasis on social and reproductive health issues. Finally, for sample sizes of clinical trials specifically, those of the most cited articles ranged from a median of about 1500 to 2500, whereas those of the least cited articles ranged from 30 to 40.

Background

Medical researchers who publish want their papers to be highly cited. Three necessary conditions for highly cited papers under control of the research author(s) are

Received January 16, 2007

Address for correspondence:
RONALD N. KOSTOFF
Office of Naval Research, Arlington, VA 22217, USA
E-mail: kostofr@onr.navy.mil

0138–9130/US $ 20.00

high intrinsic quality, high research activity discipline, and high circulation journal.[1–5] The present article examines other characteristics of highly cited medical papers, to ascertain unique attribute patterns in these papers. It extends a comparative method used by the author in recent studies, whereby the attributes of the most cited papers are compared with those of the least cited papers (e.g., Ref. 6). This contrast between most and least has been shown to delineate differences between the two groups dramatically, for single discipline studies. It was desired to extend the technique to a more heterogeneous discipline (general medical studies), and use *The Lancet*, a British-based medical journal, as the test-bed. *The Lancet* is a well-regarded leading medical journal, and contains a substantial number of very highly cited papers.

Approach

A database of *Lancet* papers published in a narrow time window (for time normalization) and accessed through the Science Citation Index (SCI) was generated, and the detailed attributes (characteristics) of most and least cited papers were identified. Specifically, all documents classified by the SCI as articles and published in *Lancet* from 1997–1999 were examined initially. The majority of records classified as Articles (~2/3) did not have Abstracts. The articles with Abstracts are viewed as complete research articles, and are analyzed in more detail. Characteristics evaluated for each of the most and least cited articles are based on the author's recent analyses of most and least cited articles in neuro-psychology[6] and desalination journals, and the anthrax discipline, with the exception that a wider range of characteristics was used for the present analysis.

Two key issues in the approach are the selection of citations as a key metric, and the quantification of citations used to define 'high' and 'low'. Citations are used as a metric for research impact/ quality, and allocation of papers into high and low citation categories is a proxy for separation into high and low quality categories. However, for clinical trials specifically, some caveats are required. As Ioannidis states,[7] in an analysis of highly cited clinical trials, "Of 49 highly cited original clinical research studies, 45 claimed that the intervention was effective. Of these, 7 (16%) were *contradicted* by subsequent studies, 7 others (16%) had found effects that were stronger than those of subsequent studies, 20 (44%) were replicated, and 11 (24%) remained largely unchallenged."

The second key issue is the quantitative determination of most and least cited. If too few articles are selected (e.g., the top or bottom two or three, or 1%), the statistics will be insufficient to allow for general conclusions. If too many articles are selected (e.g., the top or bottom 40 or 50, or 20%), the gradations within each category will be too large for 'most' and 'least' to have meaning. Based on previous studies, and on tests where the number/fraction of articles in 'most' and 'least' categories were examined,

the top and bottom 5% of articles with Abstracts were selected for each year, and their characteristics compared. The number of articles in each category was kept constant at seventeen per year.

Results

Tables 1 and 2 list detailed characteristics of the most cited and the least cited research articles for 1998. Tables 1a and 1b contain the bibliometrics, and Tables 2a and 2b contain the medical/technical issues.

In Tables 1a and 1b, starting from the second from left column, the characteristics listed are number of authors, number of references, number of citations as shown in the Times Cited field in the Science Citation Index, number of Abstract words, number of pages, reprint author country, reprint author institution type, and reprint author institution name. The reprint author is typically the first author. In Tables 2a and 2b, starting from the second from left column, the characteristics listed are medical condition/ theme, approach, study type, and sample size.

Table 3 summarizes the bibliometrics results for 1997–1999 for most and least cited articles, while Table 4 summarizes the medical issues results for the same records. The numbers in parentheses after the text entries reflect the number of occurrences. Only those entries with occurrences greater than one are shown.

Table 1a. Most cited – bibliometrics
1998 most cited articles

art #	#auth	#refs	#cites	#abs words	#pages	1st auth country	1st auth inst type	1st auth inst name
1	400	47	2799	529	17	England	HOSP	Radcliffe
2	10	33	1563	438	8	Sweden	UNIV	Uppsala
3	6	27	1232	606	17	England	HOSP	Radcliffe
4	11	27	1014	389	7	France	HOSP	Pitie Salpetriere
5	13	20	895	423	12	England	HOSP	Radcliffe
6	188	30	782	239	6	Usa	HOSP	Cleveland Clin
7	25	24	730	286	8	New Zealand	UNIV	Wellington
8	326	19	650	579	13	England	HOSP	Radcliffe
9	9	36	548	277	4	Usa	HOSP	Brigham/ Womens
10	5	32	507	323	7	Italy	HOSP	Salvatore Maugeri
11	5	8	484	328	9	Scotland	HOSP	Western Gen
12	4	25	480	166	3	Australia	HOSP	St. Vincents
13	13	19	463	262	7	Germany	UNIV	Heidelberg
14	119	39	447	296	7	Canada	HOSP	London Health
15	9	18	412	307	4	England	HOSP	Royal Marsden
16	8	20	408	341	5	Italy	HOSP	Inst Oncology
17	5	33	407	288	5	USA	HOSP	Brigham/Womens

Table 1b. Least cited – bibliometrics
1998 least cited articles

art #	#auth	#refs	#cites	#abs words	#pages	1st auth country	1st auth inst type	1st auth inst name
314	1	19	0	68	3	Scotland	UNIV	Glasgow
313	2	30	0	202	5	Netherlands	UNIV	Utrecht
312	2	37	0	80	5	Canada	UNIV	Toronto
311	2	35	2	111	3	USA	HOSP	Texas Arrythmia
310	1	13	4	124	4	England	UNIV	London
309	6	15	4	266	3	Austria	HOSP	Wilhelminenspital
308	3	4	5	198	1	England	HOSP	Charing Cross
307	1	15	5	143	5	England	FOUND	Global Forum
306	2	16	5	127	4	Norway	INST	Diakonhjemmets
305	2	6	6	154	3	USA	UNIV	Cornell
304	8	13	6	213	3	Australia	GOVT	Achs
303	3	27	6	297	5	Scotland	HOSP	Western Gen
302	3	43	8	62	5	Wales	UNIV	Cardiff
301	1	55	8	72	6	USA	UNIV	St. Louis
300	10	0	9	88	2	England	FOUND	Aids Consortium
299	6	12	9	175	3	England	UNIV	Nottingham
298	3	13	9	314	3	France	UNIV	Reims

Table 2a. Most cited – medical issues
1998 most cited articles

Art #	Medical condition/theme	Approach	Study type	Sample size	Outcome
1	Diabetes	Drugs	Clinical trial	Large (3867)	Partially favorable
2	Hypertension	Drugs	Clinical trial	Large (18790)	Favorable
3	Breast cancer	Drugs	Meta-analysis	Large (37000)	Partially favorable
4	Hepatitis virus	Drugs	Clinical trial	Large (832)	Favorable
5	Diabetes	Drugs	Clinical trial	Large (2241)	Moderately favorable
6	Coronary stent	Drugs	Clinical trial	Large (2399)	Favorable
7	Allergies	Questionnaires	Epidemiology	Large (463801)	Favorable
8	Breast cancer	Drugs	Clinical trial	Large (30000)	Moderately favorable
9	Breast cancer	Blood assays	Clinical assay	Large (1017)	Favorable
10	Myocardial infarct	Heart monitor	Clinical trial	Large (1284)	Favorable
11	Carotid stenosis	Surgical	Clinical trial	Large (3024)	Moderately favorable
12	HIV-1	Mechanism model	Theoretical	n.a.	Hypothesis
13	Stroke	Drugs	Clinical trial	Large (800)	Marginally favorable
14	Multiple sclerosis	Drugs	Clinical trial	Large (533)	Favorable
15	Breast cancer	Drugs	Clinical trial	Large (2494)	Inconclusive
16	Breast cancer	Drugs	Clinical trial	Large (5408)	Inconclusive
17	Myocardial Infarct	Blood Assays	Clinical assay	Large (14916)	Favorable

Table 2b. least cited – medical issues
1998 least cited articles

Art #	Medical condition/theme	Approach	Study type	Sample size	Outcome
314	Stroke	Treatment protocol	Assessment	n.a.	n.a.
313	Stroke	Treatment protocol	Assessment	n.a.	n.a.
312	Stroke	Treatment protocol	Assessment	n.a.	n.a.
311	Arrythmia	Treatment hypothesis	Theoretical	n.a.	n.a.
310	Global health	Evaluation	Propose reforms	n.a.	n.a.
309	Torture injuries	Bone scans	Clinical trial	small (50)	favorable
308	Pulmonary embolism	Drugs	Case study	small (1)	unfavorable
307	Who' impact	Evaluation	Propose reforms	n.a.	n.a.
306	Social impact	Evaluation	Propose reforms	n.a.	n.a.
305	Food aid	Evaluation	Assessment	n.a.	n.a.
304	Psittacosis	Epidemiology	Clinical diagnostics	small (16)	favorable
303	Brain tumors	Surgical	Clinical trial	small (40)	favorable
302	Data interpretation	Teaching	Protocols	n.a.	n.a.
301	Acid-base disorders	Diagnostic tests	Clinical diagnostics	n.a.	n.a.
300	HIV	Treatment access	Propose reforms	n.a.	n.a.
299	Placental perfusion	Diagnostic tests	Clinical trial	small (15)	favorable
298	Fertilization	Questionnaire	Survey	small (48)	inconclusive

Table 3. 1997–1999 – most/least cited – bibliometrics – summary

Art #	#auth	#refs	#cites	#abs words	#pages	1st auth country	1st auth inst type	1st auth inst name
1997 summary – most cited								
Average	46.59	30.82	692.1	358.8	7.059	W. Eur (10)	Univ (9)	Harvard (3)
Median	9	29	660	336	7	N. America (5)	Hosp (5)	
Std dev	75.96	10.68	220.2	96.33	2.703		Inst (3)	
1997 summary – least cited								
Average	3.118	21.06	5.118	178.1	3.529	W. Europe (9)	Univ (7)	
Median	2	20	5	142	3	N. America (5)	Hosp (5)	
Std dev	2.369	10.56	3.257	105	1.281		Inst (3)	
1998 summary – most cited								
Average	68	26.88	813	357.5	8.176	West Eur(11)	Univ(3)	Radcliffe(4)
Median	10	27	548	323	7	N. America(4)	Hosp(14)	Brigham/womens(2)
Std dev	122.1	9.393	606.9	121.7	4.231	Australasia(2)		
1998 summary – least cited								
Average	20.0	10.9	209.0	118.7	2.6	West Eur(12)	Univ(9)	
Median	2	15	5	143	3	N. Amer(4)	Hosp(4)	
Std dev	2.6	14.9	3.1	80.0	1.3		Found(2)	
1999 summary – most cited								
Average	32.83	19.21	342.5	207.3	4.506	West Eur (14)	Univ (9)	Uppsala (3)
Median	11	29	489	258	7	N. America (2)	Hosp (6)	
Std. dev.	40.29	8.276	323.4	110.8	2.352		Inst (2)	
1999 summary – least cited								
Average	30.51	18.88	316.4	197.5	4.26	W. Eur. (10)	Univ (12)	
Median	4	5	1	38	2	Australasia (3)	Hosp (2)	
Std. dev.	37.72	7.65	309.9	107.1	2.323	N. America (2)	Inst (3)	

Table 4. 1997–1999 – most/least cited – medical issues – summary

Art #	Medical condition/theme	Approach	Study type	Sample size	Outcome
1997 summary – most cited					
	Coronary (6)	Drugs (8)	Clin trial (8)	Large	Favorable (12)
	Mortality (3)	Data analysis (4)	Lab tests (5)	Median (1734)	Moderately fav (4)
	Breast cancer (2)	Biopsy (2)	Epidemiol (4)		
		Lab tests (2)			
1997 summary –least cited					
	Radiology (4)	Overview (5)	Overview (6)	Large (4)	Favorable (5)
		Lab tests (3)	Assessments (2)	Median (755)	
		Analysis (2)	Epidemiology (2)	Small (2)	
		Survey (2)		Median (7)	
1998 summary – most cited					
	Breast cancer(5)	Drugs(11)	Clinical trial(12)	Large(16)	Favorable (8)
	Diabetes(2)	Blood assays(2)	Clinical assays(2)	(median 3445)	Moderately favor (3)
	Myocardial infarct(2)				Partially favor(2)
1998 summary – least cited					
	Social health(4)	Evaluation(4)	Assessment(4)	Small(6)	N.a. (11)
	Stroke(3)	Treatment protocol(3)	Propose reform(4)	(median 28)	Favorable (4)
		Diagnostic tests(2)	Clinical trial(3)		Unfavorable (1)
			Clinical diagnostics(2)		Inconclusive (1)
1999 summary – most cited					
	Cardiovascular(6)	Drugs(9)	Clin trial(14)	Large(15)	Favorable (9)
	HIV-1(5)	Assessment(3)		Median(1000)	Inconclusive (4)
	Hypertension(2)				Moderately fav (3)
1999 summary – least cited					
		Lab test (6)	Case study (4)	Small (7)	Favorable (9)
		Data analysis (4)	Lab test (4)	Large (3)	Moderately favorable (2)
		Evaluation (2)	Assessment (4)		
		Survey (2)	Survey (2)		

For all three years, there are substantial differences between the most and least cited articles in both the bibliometrics and medical/technical issues. Compared to the least cited articles, the most cited article bibliometrics reflect, for the three years, approximately:

- Fifteen to twenty times the average number of authors per article, and three to five times the median number of authors per article (see Reference 8 for a network-based depiction of numbers of authors vs numbers of citations);
- Thirty three to three hundred percent greater average number of references per article, and fifty to six hundred percent greater median number of references per article (see Reference 9 for relationship of number of references to citations);
- One hundred and forty to one thousand times the average number of citations per article, and 110 to 490 times the median number of citations per article;

- Two to four times the average number of Abstract words per article, and 2.5 to almost seven times the median number of Abstract words per article;
- Two and a half to 3.5 times the average number of pages per article, and 2.5 to 3.5 times the median number of pages per article;
- There are no significant regional distribution differences between the most and least cited articles (only those regions with more than one article are listed);
- While the least cited articles tend to be predominately from universities, the most cited are mixed. The significance of this result is muted, because many of the hospitals tended to be part of a university, thereby blurring the university/ hospital distinction;
- Over the three year period, four institutions were represented three or more times in the most cited category (Radcliffe – 5; Uppsala – 4; Harvard – 3; Salpetriere – 3), while no institutions were represented three or more times in the least cited category.

The most cited articles' medical themes emphasize breast cancer, diabetes, coronary circulation, and HIV immune system problems, focusing on large scale clinical trials of drugs. The 1997 most cited articles were skewed by three publications from a large scale epidemiological study (Burden of Disease), thereby reducing the dominant fraction of clinical drug trials that characterized the other two years. The least cited articles' themes essentially do not address the above medical issues, especially from a clinical trials perspective, cover a much broader range of topics, and have much more emphasis on social and reproductive health issues. Because of the different nature of the least cited articles' studies relative to the most cited, many of the least cited articles' outcomes were non-applicable in the same sense as that of a drug clinical trial outcome. The favorability or unfavorability of results (assessed, but not included in tables) did not seem to be a major factor differentiating the most and least cited articles.

Median sample sizes of the most cited articles tend to be large, ranging from 1000 to 3500, while sample sizes of the least cited articles tend to be a couple of orders of magnitude smaller. However, 1997 did contain four of the least cited articles with a median sample size of 755, but none of these samples were from clinical drug trials. For sample sizes of clinical trials specifically, those of the most cited articles ranged from a median of about 1500 to 2500, whereas those of the least cited articles ranged from 30 to 40. References 10-11 provide further evidence that higher citation rates are correlated with larger clinical trial sample sizes, but exceptions can be found.

Discussion and conclusions

Medicine has many facets, including adequacy and affordability of health care, disease and injury prevention, public health education, lab research and clinical trials, theory and experiment, individual and global health issues, and epidemiology. Out of all these possibilities that could be of substantial interest to the medical research community, *The Lancet* readership community has chosen to emphasize high citations to large-scale clinical drug trials on breast cancer, diabetes, coronary circulation, and immune system problems, reported by many authors in long well-referenced papers, for the time period chosen.

References

1. LEHRL, S., The citation frequency for prominent researchers in German otorhinolaryngology over 10 years. *HNO*, 53 (5) (2005) 415–422.
2. SMITH, S. D., Is an article in a top journal a top article? *Financial Management*, 33 (4) (2004) 133–149.
3. LEE, K. P., SCHOTLAND, M., BACCHETTI, P., BERO, L. A., Association of journal quality indicators with methodological quality of clinical research articles. *JAMA – Journal of the American Medical Association*, 287 (21) (2002) 2805–2808.
4. KOSTOFF, R. N., The use and misuse of citation analysis in research evaluation. *Scientometrics*, 43 (1) (1998) 27–43.
5. CALLAHAM, M., WEARS, R. L., WEBER, E., Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA – Journal of the American Medical Association*, 287 (21) (2002) 2847–2850.
6. KOSTOFF, R. N., BUCHTEL, H., ANDREWS, J., PFEIL, K., The hidden structure of neuropsychology: text mining of the journal Cortex: 1991–2001. *Cortex*, 41 (2) (2005) 103–115.
7. IOANNIDIS, J. P. A., Contradicted and initially stronger effects in highly cited clinical research. *JAMA – Journal of the American Medical Association*, 294 (2) (2005) 218–228.
8. BORNER, K., DALL'ASTA, L., KE, W. M. et al., Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams. *Complexity*, 10 (4) (2005) 57–67.
9. SCHLOEGL, C., STOCK, W. G., Impact and relevance of LIS journals: A scientometric analysis of international and German-language LIS journals – Citation analysis versus reader survey. *Journal of the American Society for Information Science and Technology*, 55 (13) (2004) 1155–1168.
10. GLUUD, L. L., SORENSEN, T. I. A., GOTZSCHE, P. C. et al., The journal impact factor as a predictor of trial quality and outcomes: Cohort study of hepatobiliary randomized clinical trials. *American Journal of Gastroenterology*, 100 (11) (2005) 2431–2435.
11. EASTERBROOK, P. J., BERLIN, J. A., GOPALAN, R. et al., Publication bias in clinical research. *Lancet*, 337 (8746) (1991) 867–872.