BMJ

Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: retrospective cohort study

Cynthia Lokker, research associate, K Ann McKibbon, associate professor, R James McKinlay, data analyst, Nancy L Wilczynski, research manager, R Brian Haynes, professor

ABSTRACT

Objective To determine if citation counts at two years could be predicted for clinical articles that pass basic criteria for critical appraisal using data within three weeks of publication from external sources and an online article rating service.

Design Retrospective cohort study.

Setting Online rating service, Canada.

Participants 1274 articles from 105 journals published from January to June 2005, randomly divided into a 60:40 split to provide derivation and validation datasets.

Main outcome measures 20 article and journal features, including ratings of clinical relevance and newsworthiness, routinely collected by the McMaster online rating of evidence system, compared with citation counts at two years.

Results The derivation analysis showed that the regression equation accounted for 60% of the variation (R^2=0.60, 95% confidence interval 0.538 to 0.629). This model applied to the validation dataset gave a similar prediction (R^2=0.56, 0.476 to 0.596, shrinkage 0.04; shrinkage measures how well the derived equation matches data from the validation dataset). Cited articles in the top half and top third were predicted with 83% and 61% sensitivity and 72% and 82% specificity. Higher citations were predicted by indexing in numerous databases; number of authors; abstraction in synoptic journals; clinical relevance scores; number of cited references; and original, multicentred, and therapy articles from journals with a greater proportion of articles abstracted.

Conclusion Citation counts can be reliably predicted at two years using data within three weeks of publication.

INTRODUCTION

Evidence based medicine incorporates the best clinical evidence with clinicians' expertise and patients' values in clinical decision making. Finding and disseminating such evidence is a difficult task given the amount of research being published continuously. The task is made more difficult by high quality and clinically relevant healthcare articles being diluted in the larger body of articles reporting basic science, opinions, news, and lesser quality studies.

If the importance of an article to clinical practice could be predicted soon after publication, then a focused push of such articles could be made to clinicians and other readers who could potentially use them. The articles could also be quickly sent to authors and publishers of information resources, systematic reviews, clinical practice guidelines, and educational programmes. Various factors have been used to predict the number of citations for an article, such as methodological quality,[1] a journal's science citation index impact factor,[2][3] and media coverage.[4] The quality[5] and accuracy[6] of web pages are also associated with the number of web citations—that is, links to that site from other sites. The number of times an article is cited in subsequent publications is an attractive measure of importance, or at least notice, by peers and others as it is readily available, but it has no applicability until citation counts are accrued in the years after publication, peaking at around three years.[7]

Studies have shown that several attributes of articles are associated with higher citation counts. One study looked at indicators of quality for predicting citation counts in emergency medicine and reported a pseudo R^2 of 0.14 (14% of the variance), with impact factor as the only significant variable and presence of a control group, subjective newsworthiness, and sample size as the next most important factors after adjustment for impact factor.[2] Quality, as represented by clear reporting of the research question, primary outcome, and appropriateness of data analysis, was found not to influence citation counts in psychiatric journals, whereas an unclear description of statistical analysis reduced the number of citations.[8] Another study found that articles published in three of the top medical journals that declared industry funding and industry favoured results had group authorship and increased sample size, and those relating to oncology and cardiovascular medicine were associated with higher annual citation rates.[9] Thirty three per cent of the variance in citation counts of *BMJ* articles were found to be based on counts of online hits and number of

pages.[10] Other factors influencing citation counts include nationality,[11 12] number of authors,[13 14] number of pages,[13] methods,[11 15] reviews,[16] therapy,[16] and online (open access) availability.[17] Some authors have even shown that men and those with surnames in the first half of the alphabet obtain more citations.[18]

We determined how well the citation count of articles that pass basic critical appraisal criteria for healthcare research can be predicted at two years using data available at the time of publication or within three weeks thereafter. We included article ratings that are now routinely collected through the McMaster online rating of evidence system[19] but that have not yet been studied for this purpose.

METHODS

The Health Information Research Unit at McMaster University identifies newly published, high quality healthcare evidence (studies and systematic reviews) and brings them to the attention of practising clinicians through several print and electronic media. The group is involved in the production of the evidence synopsis journals *ACP Journal Club, Evidence-Based Medicine* and *Evidence-Based Nursing*, and email alerting services bmjupdates+ and Medscape Best Evidence Alerts. The group is also involved in the production of such information resources as Physicians' Information and Education Resource, BMJ Clinical Evidence, and McGraw-Hill's Harrison's Practice. The identification process starts with the reading of over 130 journal titles.[20] Trained research associates apply explicit reading criteria to all articles ensuring methodological rigour of the original articles and reviews selected.[20]

Articles that pass the filters for methods and content set by the Health Information Research Unit are categorised by research staff as being important to one or more of 59 clinical disciplines (for example, family practice, internal medicine and its subspecialties, paediatrics, and surgery). Each article is sent for rating to practitioners for each pertinent discipline through the McMaster online rating of evidence system[19] for rating of the article's importance (clinical relevance and newsworthiness). This system uses the services of over 4000 practitioners, about half from primary care and the rest from other specialties. The article is rated by at least three practitioners in each of the article's clinical domains. Clinical relevance is rated using a 7 point Likert scale ranging from 1 "not relevant" to 7 "directly and highly relevant." Newsworthiness is rated from 1 "not of direct clinical interest" to 7 "useful information, most practitioners in my discipline definitely don't know this." Articles with minimum ratings of 4 for each scale are sent as email alerts to subscribers, and a subset of the more important ones are selected for summary in the three evidence based synoptic journals. All articles published from January to June 2005 that passed our methods filter and had a minimum average clinical rating of 4 for each scale for at least one clinical discipline formed the basis for our study. This period was sufficient to allow for the accumulation of citations for about two years.

The attributes we chose a priori for our analysis were readily available within three weeks of publication of the article. These aspects of the article and journal had either been studied previously or were ones that we considered could be predictive of the number of citations for the article. We collected 17 article specific and three journal specific variables for each article. These are listed in table 1 along with our hypotheses of how each variable would possibly affect the number of citations.

Citation counts were collected from the Institute of Scientific Information web of science between 17 April and 8 June 2007, giving almost 24 months of citation accrual per article. We did not look at self citation[21] or the sex of the first author[18] as these data are difficult to ascertain. We also did not use alphabetical positioning of the first author's surname,[18] number of online hits,[10] or downloads of articles[22] as they have not been consistently shown to affect citation rates or are difficult to ascertain.

Analysis

We carried out multiple regression using a random split of our sample into 60:40 derivation-validation datasets. We used Stata Intercooled 9.0 software. All 20 variables were included in the analysis. After the regression we tested colinearity with the variance inflation factor, which formed the basis for removal of colinear variables. Variance inflation factor values under 5 are considered to not influence the outcome of the regression, over 10 has a strong effect on the regression.[23] We also tested for outliers using added

Table 1 | Multiple regression results from analysis of 757 articles in derivation dataset*

Independent variable	Regression coefficient (95% CI)	P value
No of databases indexed	0.039 (0.032 to 0.045)	<0.001
No of authors	0.087 (0.063 to 0.112)	<0.001
Proportion of articles abstracted in 2005	8.18 (4.35 to 12.0)	<0.001
Abstracted in evidence based medicine journals	0.839 (0.422 to 1.26)	<0.001
Mean clinical relevance score	0.418 (0.203 to 0.633)	<0.001
No of pages	−0.011 (−0.017 to −0.005)	<0.001
Structured abstract	−0.808 (−1.31 to −0.308)	0.002
No of cited references	0.004 (0.001 to 0.008)	0.008
Original v review article	0.477 (0.117 to 0.837)	0.009
Multicentre study	0.367 (0.073 to 0.660)	0.014
Therapy v other	0.339 (0.047 to 0.631)	0.023
Mean newsworthiness score	0.133 (−0.053 to 0.320)	0.161
No of participants	<0.001	0.295
No of disciplines rating article	0.038 (−0.045 to 0.120)	0.371
Time to rating article	−0.009 (−0.037 to 0.019)	0.513
Word count of abstract	−0.0003 (−0.002 to 0.001)	0.658
Semistructured abstract	0.071 (−0.357 to 0.498)	0.746
Nation of first author	−0.037 (−0.275 to 0.202)	0.762
No of views or alerts sent	−0.069 (−1.82 to 1.68)	0.938
_constant	−3.02 (−4.23 to −1.80)	<0.001

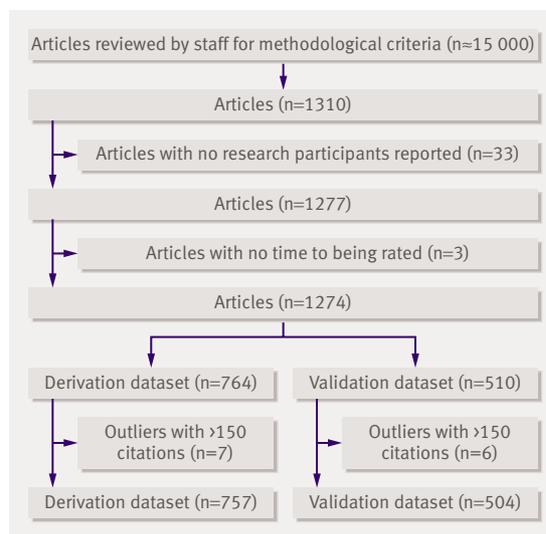*$F_{19,737}$=58.12, P<0.001, R^2=0.60 (95% confidence interval 0.538 to 0.629).

Fig 1 | Flow of study data

value plots and for the normality of the regression residuals using residual plots.

To determine if our model has a good chance of fitting other samples we carried out a cross validation.[24][25] Using the regression model from the derivation dataset we predicted values for each case in the validation dataset and computed the R^2 values. Subtracting the R^2 generated by applying the derivation model to the validation dataset from the derivation R^2 value provided the "shrinkage on cross validation" of the analysis—a measure of how well the derived regression equation matched the actual data from the validation dataset.[24][25] We further used the derivation dataset to determine the sensitivity and specificity of our model in detecting the articles with citations greater than the median and those that were in the top third of cited articles.

The original regression indicated several outlying cases. These were determined by looking at the added variable plots for each variable. These plots show which cases are exerting disproportionate influence on the regression model.[23] The cases consistently uncovered by the added variable plots were articles with extremely high citation counts. Residual plots indicated some non-normality in the model. Because citation counts were highly negatively skewed we transformed them with square roots of the counts. The variance inflation factor indicated that the proportion of articles that passed reading criteria per journal in 2005 correlated with the proportion that was abstracted, and we removed the proportion that passed the reading criteria from the analyses.

RESULTS

Originally about 15 000 articles were reviewed by the journal production staff from January to June 2005. Of those, the sample retrieved from the McMaster online rating of evidence system contained 1310 articles that passed the criteria for methods and had average ratings of at least 4 (out of 7) for each of clinical relevance and

newsworthiness. Forty nine articles were excluded: 33 had no details on the number of participants, three had no time to being rated values, and 13 had citation counts greater than 150 (which were removed from the regression only). These 13 articles included 11 from the *New England Journal of Medicine* and one each from the *Lancet* and the *Annals of Internal Medicine.* The final analysis was carried out on 1261 articles published in 105 journals. Figure 1 shows the flow of the study data.

Citation counts for the derivation dataset (n=757) varied from 0 to 128, the mean was 13.1 citations, and the median was 6, with substantial negative skew; 166 articles had no citations. Cochrane reviews and articles from the Health Technology Assessment (HTA) database accounted for 24% of the sample (n=182).

The resulting multiple regression on the derivation dataset was highly significant (table 1): R^2=0.60 (95% confidence interval 0.538 to 0.629), $P<0.001$. Nine article specific variables were statistically significant: the number of authors, selection for abstraction in a synoptic journal, clinical relevance score, number of pages, structured abstract, number of cited references, original article, multicentred study, and study about therapy. Two journal specific variables—the number of bibliographic databases in which the journal was indexed and the proportion of the published articles that were abstracted from the journal in 2005—also predicted higher citations.

The hypothesis was that number of pages and a structured abstract would have a positive influence and
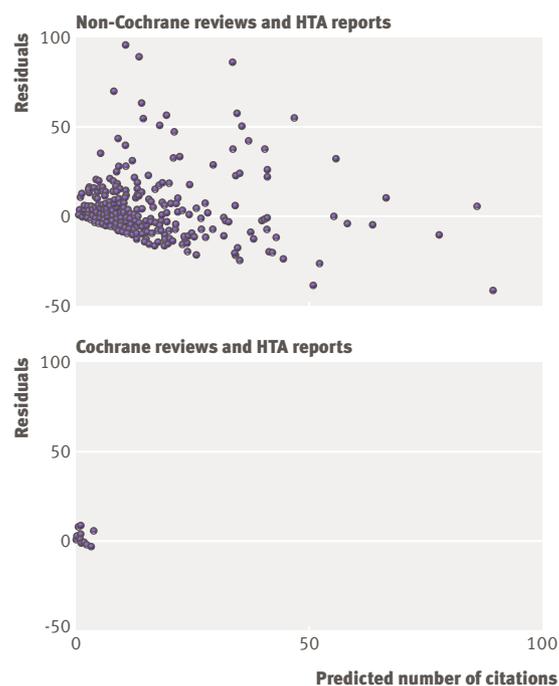

Fig 2 | Plot of residuals versus predicted citations for 504 articles from validation dataset (outliers excluded). Perfect agreement is indicated by observations along zero line on vertical axis. Top graph represents 396 articles in validation dataset, excluding 108 articles that are Cochrane reviews or reports from the Health Technology Assessment (HTA) database (bottom graph)

that being an original study would have a negative influence (table 2). Each of these three variables had statistically significant counterintuitive results possibly driven by the number of Cochrane reviews in the dataset (23% in the derivation dataset, 20% in the validation dataset). These reviews tend to be much longer than journal articles and to have structured abstracts. They also had low citation rates, with a mean of 0.46 (range 0-9) citations.

To test how well the model fitted another dataset, the derivation regression model was used to create predicted values for each case in the validation dataset. For this validation dataset (n=504) citation counts varied from 0 to 120, the mean was 12.9 citations, and the median was 7, with substantial negative skew; 109 articles had no citations. Cochrane reviews and HTA reports accounted for 21% of the sample (n=108).

The R^2 based on the validation dataset was 0.56 (95% confidence interval 0.48 to 0.60, P<0.001). The shrinkage on cross validation between the two datasets was low at 0.04, or 4%. A shrinkage that is below 0.10 suggests that the model can be generalised to other datasets[24] containing similar articles that pass the
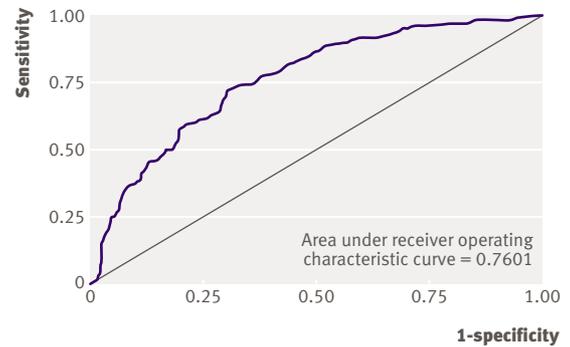

Fig 3 | Area under receiver operating characteristic curve for regression model in predicting citation rates at two years. Threshold was based on articles with a median of more than seven citations

critical appraisal screen. A plot of the residuals shows the distribution of the difference in observed and predicted citation counts (fig 2). The average difference between observed and expected counts was 4.95 (95% confidence interval 2.03 to 7.86).

Using the validation results with the six outliers included, the sensitivity and specificity of the model (derivation dataset) in predicting citation counts (validation dataset) at cut points representing the top half (>7 citations) and the top third (>12 citations) of cited articles was determined. The model had a sensitivity of 83.3% and a specificity of 71.5% for predicting the top half of cited articles and a sensitivity of 66.1% and a specificity of 82.2% for predicting the top third. The area under the receiver operating characteristic curve was 0.76 (95% confidence interval 0.722 to 0.80) for a threshold set at seven articles—the median (fig 3). The model did fairly well at discriminating the top performing articles.

DISCUSSION

The number of citations an article accrues by two years after publication can be predicted with about 60% surety using data within three weeks of publication. The ability to predict citation counts in this study is higher than others have reported. One study was only able to predict 14% of the variance for the annual citation rates using papers on emergency medicine and 12 possible predictor variables,[2] whereas another study predicted 20% using articles from *JAMA,* the *Lancet,* and the *New England Journal of Medicine.*[9] Our sample, however, represents a select group of articles that passed methodological criteria and the greater predictability could result from lower variability in the dataset.

We have also shown that physician rated clinical relevance at the time of publication is related to citation counts at two years. This replicates results from a study which showed that the citation counts of papers important to rhinology were highly related to clinical utility.[26]

The quality of studies has been shown to be weakly or moderately related to citation counts.[11] In our study we could not assess the influence of quality of research

Table 2 | Variables to predict citation counts at two years and hypothesised influence on increased citation counts

Predictor variables	Hypothesised influences
Article specific from external sources:	
No of authors	More authors
Residence of first author in North America	North America
No of pages	Longer article
No of references in bibliography	More references
No of participants	More participants
Structured abstract*	Structured abstracts
Length of abstract	Longer
Multicentre studies	If multicentred
Original article rather than systematic review	If systematic review
Dealing with therapy	If therapy
Article specific from internal sources:	
No of disciplines chosen relevant to article (article's breadth of interest)†	More disciplines
Average relevance scores over all raters†	Higher scores
Average newsworthiness scores over all raters†	Higher scores
Average time taken by raters to rate article†	More time
Whether article was selected for abstraction in 1 of 3 synoptic journals‡	If yes
No of views per email alert sent§	More views per alert
Journal specific using internal data:	
Proportion of articles that passed criteria (2005)‡	Higher proportion
Proportion abstracted by 3 synoptic journals‡¶	Higher proportion
Journal specific using external data:	
No of databases that index journal**††	More databases

*Eight headings as proposed by Haynes et al[20]; semistructured with general introduction, methods, results, discussion or conclusion (IMRAD) format; or unstructured.
†From McMaster online rating of evidence system.
‡From Health Information Research Unit.
§From bmjupdates+ (http://bmjupdates.mcmaster.ca).
¶Calculated as number of articles abstracted per journal in 2005 divided by total number of articles published.
**Count based on list of active "abstracting and indexing sources" listed in Ulrich's periodicals directory (www.ulrichsweb.com) for each journal.
††From Ulrich's periodicals directory.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Citation counts are markers of an article's importance but are not available for months after publication

Research shows that various attributes of an article are related to higher citation rates, but the predictive value of these factors is limited

WHAT THIS STUDY ADDS

Features of methodologically sound articles predicted citation counts with higher reliability than previously found

Ratings of clinical relevance by practising clinicians are significantly associated with citation counts at two years

methods because we only included articles that had passed basic methodological criteria, and we had not graded quality within this higher quality dataset.

Several groups have predicted citations using journal impact factors. When impact factors from 2004 were included in our derivation regression the sample size was reduced by 182 articles because Cochrane reviews and HTA reports do not have impact factors. The resulting regression on the smaller dataset gave an R^2 of 0.47 (95% confidence interval 0.39 to 0.51). Given the importance of Cochrane reviews and HTA reports we chose to include them and not to use the impact factor as a predictor variable.

Cochrane reviews and HTA reports are systematic reviews of evidence, which are typically lengthy and have structured abstracts. Their average citation rates were low compared with the journal articles in our study. We believe that this underlies our findings of negative correlations between citation counts and number of pages, the presence of a structured abstract, and review articles. Without Cochrane reviews and HTA reports in our analysis, comparisons between the number of pages and original article versus review articles were no longer statistically significant, although structured abstracts remained negatively correlated with citation counts. As we had no prior hypothesis on the influence of Cochrane reviews and HTA reports on citation counts, however, we kept them in our regression model. We are aware that the reduced variability among these publications, as seen by the clustering of the residuals in figure 2, improved the performance of our regression.

Predicting citation counts early could allow providers of information resources to quickly identify those articles that are likely to have an impact on clinical practice. We hope to use this model to refine our approach to "pushing" detected articles to practising clinicians, authors, and publishers.

The positive association that we found between clinical ratings and citation counts could be because early dissemination of such articles actually leads to higher citation rates, not merely predicts citation. We could not assess this possibility, but the intended targets of the dissemination process are practising clinicians, not scientists who write papers. Thus our findings support an association that constitutes "criterion validity," in that ratings predict an accepted measure of research merit.

Strengths and limitations

The strengths of this study include the large number of included articles (n=1261), the magnitude of the association in the derivation dataset, and the agreement between the results from the derivation and validation datasets. We have shown a statistically significant relation between ratings of the clinical relevance of an article and its citation count.

Several weaknesses are present in our study. Our journal subset included only 105 of the most important clinical journals, a relatively small proportion of all such journals. Therefore our results may not be readily transferable to articles in less important clinical journals or basic science articles or journals. Also, our selected articles were limited to those clinical articles that passed basic criteria for critical appraisal and they represent a small proportion of articles published in any given journal. Such a select sample would have reduced variability resulting in greater predictability. The inclusion of the Cochrane reviews and HTA reports also reduced the variability and led to greater predictability.

Conclusions

We collected data on 20 journal specific and article specific characteristics of 1261 articles from 105 clinical journals to determine if we could predict citation counts at two years. Eleven remained statistically significant in our regression model. Therefore we can predict citation counts of methodologically sound clinical studies and review articles at two years with surety using data available within three weeks after publication.

1   Lee KP. Association of journal quality indicators with methodological quality of clinical research articles. *JAMA* 2002;287:2805-8.
2   Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA* 2002;287:2847-50.
3   Barbui C, Cipriani A, Malvini L, Tansella M. Validity of the impact factor of journals as a measure of randomized controlled trial quality. *J Clin Psychol* 2006;67:37-40.

4 Stryker JE. Reporting medical information: effects of press releases and newsworthiness on medical journal articles' visibility in the news media. *Prev Med* 2002;35:519-30.

5 Griffiths KM, Tang TT, Hawking D, Chistensen H. Automated assessment of the quality of depression websites. *J Med Internet Res* 2005;7:e59.

6 Fricke M, Fallis D, Jones M, Luszko GM. Consumer health information on the internet about carpal tunnel syndrome: indicators of accuracy. *Am J Med* 2005;118:168-74.

7 Adams AB, Simonson D. Publication, citations, and impact factors of leading investigators in critical care medicine. *Respir Care* 2004;49:276-81.

8 Nieminen P, Carpenter J, Rucker G, Schumacher M. The relationship between quality of research and citation frequency. *BMC Med Res Methodol* 2006;6:42.

9 Kulkarni AV, Busse JW, Shams I. Characteristics associated with citation rate of the medical literature. *PLoS ONE* 2007;2:e403.

10 Perneger TV. Relation between online "hit counts" and subsequent citations: prospective study of research papers in the BMJ. *BMJ* 2004;329:546-7.

11 West R, McIlwaine A. What do citation counts count for in the field of addiction? An empirical evaluation of citation counts and their link with peer ratings of quality.[see comment].*Addiction* 2002;87:501-4.

12 Campbell FM. National bias: a comparison of citation practices by health professionals. *Bull Med Libr Assoc* 1990;78:376-82.

13 Bornmann L, Hans-Dieter D. Multiple publication on a single research study: does it pay? The influence of number of research articles on total citation counts in biomedicine. *J Am Soc Inf Sci* 2007;58:1100-7.

14 Figg WD, Dunn L, Liewehr DJ, Steinberg SM, Thurman PW, Barrett JC, et al. Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy* 2006;26:759-67.

15 Patsopoulos NA, Analatos AA, Ioannidis JP. Relative citation impact of various study designs in the health sciences. *JAMA* 2005;293:2362-6.

16 Bhandari M, Busse JW, Devereaux PJ, Montori VM, Swiontkowski M, Tornetta P, et al. Factors associated with citation rates in the orthopedic literature. *Can J Surg* 2007;50:119-23.

17 Eysenbach G. Citation advantage of open access articles. *PLoS Biol* 2006;4:e157.

18 Leimy R, Koricheva J. What determines the citation frequency of ecological papers? *Trends Ecol Evol* 2005;20:28-32.

19 Haynes RB, Cotoi C, Holland J, Walters L, Wilczynski NL, Jedraszewski D, et al. A second order of peer review: a system to provide peer review of the medical literature for clinical practitioners. *JAMA* 2006;2006:1801-8.

20 Wilczynski NL, McKibbon KA, Haynes RB. Enhancing retrieval of best evidence for health care from bibliographic databases: calibration of the hand search of the literature. *Medinfo* 2001;10(Pt 1):390-3.

21 Motamed M, Mehta D, Basavaraj S, Fuad F. Self citations and impact factors in otolaryngology journals. *Clin Otolaryngol* 2002;27:318-20.

22 Coats AJS. Top of the charts: download versus citations in the International Journal of Cardiology. *Int J Cardiol* 2005;105:123-5.

23 Hamilton LC. *Statistic with STATA, updated for version 9*. Belmont, CA: Thomson Books/Cole, 2006.

24 Kleinbaum DGK, Kupper LL, Muller KE. *Applied regression analysis and other multivariable methods*. 2nd ed. Boston: PWS-Kent Publishing, 1988.

25 Osborne JW. Prediction in multiple regression. *Pract Assess Res Eval* 2000;7(2).

26 Fenton JE, O'Connor A, Ullah I, Ahmed I, Shikh M. Do citation classics in rhinology reflect utility rather than quality. *Rhinology* 2005;43:221-4.