



# The impact of article titles on citation hits: an analysis of general and specialist medical journals

Thomas S Jacques<sup>1,2</sup> • Neil J Sebire<sup>2,3</sup>

<sup>1</sup>Neural Development Unit, UCL – Institute of Child Health, London, UK

<sup>2</sup>Department of Histopathology, Great Ormond Street Hospital, London, UK

<sup>3</sup>Molecular Haematology and Cancer Biology Units, UCL – Institute of Child Health, London, UK

Correspondence to: Neil J Sebire. E-mail: sebirn@gosh.nhs.uk

## DECLARATIONS

### Competing interests

None declared

### Funding

None

### Ethical approval

Not applicable

### Guarantor

NJS

### Contributorship

Both authors  
contributed equally

### Acknowledgements

None

### Reviewer

Christopher Martin

## Summary

**Objectives** Most published articles are not cited and citation rates depend on many variables. We hypothesized that specific features of journal titles may be related to citation rates.

**Design** We reviewed the title characteristics of the 25 most cited articles and the 25 least cited articles published in 2005 in general and specialist medical journals including the *Lancet*, *BMJ* and *Journal of Clinical Pathology*. The title length and construction were correlated to the number of times the papers have been cited to May 2009.

**Setting** Retrospective review of a scientific database.

**Participants** None.

**Main outcome measures** Citation rate.

**Results** The number of citations was positively correlated with the length of the title, the presence of a colon in the title and the presence of an acronym. Factors that predicted poor citation included reference to a specific country in the title.

**Conclusions** These data suggest that the construction of an article title has a significant impact on frequently the paper is cited. We hypothesize that this may be related to the way electronic searches of the literature are undertaken.

## Introduction

The number of peer-reviewed scientific publications in the medical literature is continuing to increase.<sup>1</sup> It has previously been reported that across the spectrum of science and social science specialties, more than half of all published articles are never cited by any other paper; this figure being slightly lower for biomedicine but more than 95% for some other disciplines.<sup>2</sup> Clearly,

the citation rate of an article depends on a large number of factors including the significance and availability of the journal in which it is published, publication type, its subject, and many other factors. Nevertheless, it has also previously been reported that in some disciplines, law for example, the chance of acceptance of a scientific article by a journal and its citation rate can be mathematically related to certain variables, such as the length of the title and gender of the

authors.<sup>3</sup> We hypothesized that similar features might be present in both general and specialist medical literature.

## Methods

Using Thompson ISI,<sup>4</sup> accessed May 2009, searches were performed of the Web of Science Database for publications in 2005. The journal title fields included three separate searches for the *Lancet*, *British Medical Journal* and *Journal of Clinical Pathology*, to represent general and specialist clinical medical journals, respectively. For each journal's publications during 2005, lists of the 25 most cited and 25 least cited articles were generated and the titles and number of cites data extracted. Only full articles were included; the study specifically did not include letters, editorial material or single case reports but included primary articles, review articles and meta-analyses. This was to allow comparison of title characteristics for similar article types; the aim of this study was not to compare citation rates for different article types. Title word counts, data regarding title structure and specific title words appearing most frequently in each group were generated and the frequency of these variables compared between groups using Mann-Whitney U test and comparison of proportion tests as appropriate. Power calculation indicated that 25 cases per group would reliably exclude a difference in distribution of word counts between groups of 30% at  $\alpha = 0.05$  with a power of 80%.

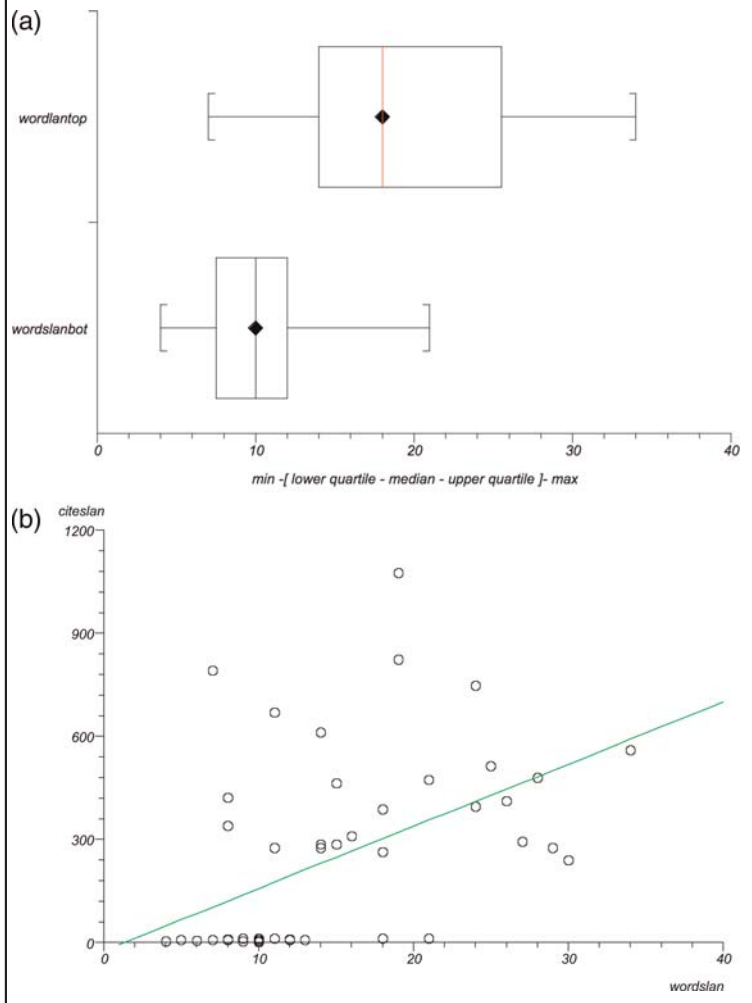
## Results

The top 25 articles in the *Lancet* (impact factor [IF] = 28.6 [2007]) were associated with 240–1075 (median 412) citations compared to 2–12 (median 8) for the lowest cited articles during the same period. The most cited articles were associated with a significantly greater number of words in the title (7–34 [median 18] compared to 4–21 [median 9]); median difference 9 words [95% CI 5–13];  $P < 0.0001$ ; Figure 1a). There was also a significant positive correlation between number of title words and citation rate across this population ( $\rho = 0.62$ , 2-sided  $P < 0.0001$ ; Figure 1b) In addition, acronyms were present in significantly more of the top cited articles, being

present in around one-third of titles, compared to the least cited, in which they were not used at all (9 of 25 [36%] versus 0 of 25 [0%]  $Z = 3.3$ ,  $P < 0.001$ ). The structure of the title was also significantly different between the groups. Specifically, a title with two components separated by a colon was significantly more common in the well-cited group, being present in more than 70% of titles compared to a minority of titles in the poorly-cited group (18 of 25 versus 10 of 25, respectively;  $Z = 2.3$ ,  $P = 0.02$ ). Finally, specific words were more common in the well-cited versus the poorly-cited groups. Words most often present in well-cited articles included: trial (13), randomized (8), study (5), cancer (4), survival (3) and risk (3), whereas in the poorly-cited group the most commonly encountered words were child(ren)/infant (5), trial (2), randomized (2), health (2), death (2) and industry (2). Most dramatically, specific country names were present in nine of the 25 most poorly-cited articles compared to none of the well-cited publications (9/25 versus 0/25;  $Z = 3.3$ ,  $P < 0.001$ ).

Similar observations characterized citation patterns in the *British Medical Journal* (IF = 9.7 [2007]). The top 25 articles had a citation range of 71–191 (median 96) compared with 1–5 (median 3) for the bottom 25. The most cited articles had significantly more title words (9–28 words; median 16) compared with the least cited articles (2–19 words, median 13;  $p = 0.005$ ). There was a significant positive correlation between the number of words in the title and the number of citations ( $\rho = 0.397$ ,  $p = 0.004$ ). Papers in the bottom 25 were also significantly more likely to refer to a specific geographic region in the title than the top 25 (10 of 25 [40%] compared with 2 of 25 [8%], respectively;  $p = 0.02$ ). The proportion of titles that were punctuated by a colon or as two sentences was also greater in the top 25 (22 of 25, 88%) than the bottom 25 (19 of 25, 76%) but this difference did not reach statistical significance. In contrast to the findings from the *Lancet*, acronyms were rare in both groups in this journal. However, the only paper with an acronym was in the highly-cited group. Specific words most commonly used in the most cited articles included: risks (13), review (8), randomized (7), controlled (7), patients (5), meta-analysis (5), cohort (5), population (3), people (3), obesity (3), doses (3), clinical (3), case-control (3) and cancer (3). The words most

**Figure 1**  
**(a) Box-whisker plot demonstrating the significantly increased number of title words in highly-cited versus poorly-cited articles in the *Lancet*. (b) Scatter plot demonstrating positive correlation between length of article title and number of subsequent citations**



commonly used in the least cited articles included: study (9), survey (6), registration (3), questionnaire (3) and national (3).

Finally, analysis was performed for a specialist journal, the *Journal of Clinical Pathology* (IF = 2.4 [2007]), and similar findings were encountered. The top 25 articles in *J Clin Path* were associated with 17–64 (median 26) citations compared to 0–1 (median 0) for the lowest cited articles during the same period. The most cited articles were associated with slightly more words in

the title (4–24 [median 12] compared to 4–21 [median 10]; median difference 2 words; NS). Acronyms were present in significantly more of the top cited articles, being present in more than one half of titles, compared to the least-cited, in which they were rarely used (13 of 25 [52%] versus 4 of 25 [16%],  $Z = 2.7$ ,  $P = 0.007$ ). A title with two components separated by a colon was present in more well-cited articles (5 of 25 versus 2 of 25, respectively;  $Z = 1.2$ ,  $P = 0.2$ ). Finally, specific words most often present in well-cited articles included: cancer (13), expression (9), prognosis/prognostic (4), immunostain\* (4) and pathogenesis (2); in the poorly-cited group the most commonly encountered words were: cancer (6), assay (3), development (3), autopsy/postmortem (2) and children/childhood (2). Specific country names were present in two of the 25 most poorly cited articles but none of the well-cited publications.

## Discussion

The results of this study have confirmed the hypothesis that certain features related to the title of a scientific article appear to be related to the number of subsequent citations it receives, these data providing potentially useful guidelines for authors of future papers.

In general, there was a strong association between increasing title length and citation rate, with the highest-scoring articles having more than twice as many words in the title than the lowest-cited articles. There are potential confounders that may have contributed to these observations. For example, we have considered the possibility that the characteristics of the title relate to particular fields, e.g. papers in highly-cited fields may also tend to have specific title characteristics. As judged by certain title words, some topics would appear more common in highly-cited papers than poorly-cited papers, e.g. the word 'cancer' appears frequently in highly-cited papers in the *Lancet* whereas the word 'children' appears more frequently in poorly-cited papers. However, it seems unlikely that the content of the paper is related in a consistent way to characteristics such as title length. Indeed, the impact of the title occurs in both generalist and specialist journals, which would

suggest that some of the effects of title construction are unrelated to the topic of the paper. Almost all literature searches are now carried out by electronic means based on searches of online databases such as PubMed or Web of Science, with many searches restricted to title or keywords only. Therefore, a longer, more comprehensive title is both more likely to contain any given search term and, therefore, be identified, and also if the title provides a clear description of the study or its findings, is also more likely to be identified as relevant on the initial search screening process. Additionally, two further factors appear significantly more likely in well-cited articles; the use of acronyms in the title and the title structure being composed of two phrases or sentences. Again, part of this relationship is likely to be due to confounding factors, such as acronyms being associated with large multicentre trials, but it is also likely that the presence of the acronym itself is beneficial for article recognition in searches, since many researchers may more commonly know or use the acronym than the full name of the gene, product or trial. This is highlighted by a previous study which reviewed all articles that included reference to the National Institute for Health and Clinical Excellence (NICE), in their titles, and suggested that such well-chosen names may increase the recognizability of public health organizations and help to communicate their roles.<sup>5</sup>

Specific words encountered within well-cited and poorly-cited article titles also provide some insight into the process, which may be used for identifying studies. Randomized, trial, cancer, survival, expression and prognosis are more commonly encountered in the well-cited group, whereas words such as children, death, autopsy, industry and development are more common in the poorly-cited group. This is likely to be in part because large randomized trials generally provide the best level of evidence for an intervention or treatment so are more likely to be cited, but it also appears that studies perceived as being related to death, specific groups such as children, or development of a technique may suggest that the findings will not be applicable to other studies. The most striking illustration of the effects of title words, however, appears to be the dramatic adverse effect of having a specific country mentioned in the title, this being present in more than one-third of poorly-cited *Lancet* articles but none of the well-

cited articles, despite many of the well-cited group representing specific studies performed in single countries. It is likely that when searching for evidence, many researchers may discount information, which is perceived to only relate to another specific country.

Our results demonstrate the massive variation in citation rates for similar article types even within major journals with high IFs.<sup>6</sup> For example, in the *Lancet*, for publications in 2005, the most cited primary research paper was cited >1000 times, whereas the least cited primary full paper was cited only three times during the same period, despite both being published in the same journal in the same year. The pattern of discordance in citation rates between well- and poorly-cited articles holds true across specialist journals as well, the main difference being that the absolute citation rate is very much lower for specialist journals. The huge difference in citation rates for articles within a given journal has been previously well-recognized, and it has been suggested that more than 50% of all citations contributing to the IF of a journal are generated by 15% of less of the articles, with some very highly-cited articles, but the majority remaining relatively poorly-cited.<sup>6</sup> Furthermore, for any given author there is a poor relationship between the IF of the journal in which an article is published and its future citation rate; simply publishing in a high IF journal does not guarantee subsequent citations.<sup>6</sup>

While we are not suggesting that changing the title of an article in isolation will lead to more citations, in practice the importance of article titles is increasing for the following reasons. It is well-established that certain specific aspects of an item, advertisement or other object, including the specific words used to describe it, can have a marked effect on its uptake.<sup>7</sup> Furthermore, given that scientific literature searches can now be performed electronically by anyone with access to the Internet via sites such as PubMed or Google Scholar, often using title searches, it is highly likely that the inclusion of these specific title modifications may lead to a greater likelihood that the article is identified and read, and hence cited. This is especially important since it has been previously suggested that many article titles may presently be inaccurate or misleading,<sup>8</sup> and a review of dermatology journals reported that the majority of articles did not report the study design used in

the title or abstract.<sup>9</sup> Furthermore, a study which examined specifically the amount of information that needs to be provided by an information retrieval system to assist healthcare practitioners in identifying clinically relevant information and evaluating its strength of evidence reported that while around 90% of article titles are informative enough to classify the publication as clinical and in the area of interest, most are insufficient for further classification of research quality. The authors note that features such as specific title content will become increasingly important with the increasing future use of computerized information retrieval systems supporting small, low-bandwidth handheld computers for use in healthcare settings.<sup>10</sup>

## References

- 1 Stringer MJ, Sales-Pardo M, Nunes Amaral LA. Effectiveness of Journal Ranking Schemes as a Tool for Locating Information. *PLoS ONE* 2008;3:e1683; doi:10.1371/journal.pone.0001683
- 2 Hamilton DP. Research papers: who's uncited now? *Science* 1991;251:25
- 3 Ayres I. *Super Crunchers: how anything can be predicted*. London: John Murray Publishers; 2008
- 4 ISI journal citation reports. See <http://admin-apps.isiknowledge.com/JCR/JCR> (last checked 17 April 2009)
- 5 Morrison DS, Batty GD. The advantages of being called NICE: a systematic review of journal article titles using the acronym for the National Institute for Health and Clinical Excellence. *J Public Health* 2009;31:127–30
- 6 Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ* 1997;314:497
- 7 Teng J. Top seven proven words that your ad copy can't live without. 2006. See [http://www.streetdirectory.com/travel\\_guide/29328/advertising/top\\_7\\_proven\\_words\\_that\\_your\\_ad\\_copy\\_cant\\_live\\_without.html](http://www.streetdirectory.com/travel_guide/29328/advertising/top_7_proven_words_that_your_ad_copy_cant_live_without.html)
- 8 Smith J, Dunstone M, Elliott-Rudder M. Health professional knowledge of breastfeeding: are the health risks of infant formula feeding accurately conveyed by the titles and abstracts of journal articles? *J Hum Lact* 2009;25:350–8
- 9 Ubriani R, Smith N, Katz KA. Reporting of study design in titles and abstracts of articles published in clinically oriented dermatology journals. *Br J Dermatol* 2007;156:557–9
- 10 Demner-Fushman D, Hauser S, Thoma G. The role of title, metadata and abstract in identifying clinically relevant journal articles. *AMIA Annu Symp Proc* 2005:191–5

© 2010 Royal Society of Medicine Press

This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/2.0/>), which permits non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.