Relation between online “hit counts” and subsequent citations: prospective study of research papers in the BMJ

Thomas V Perneger

Evaluation of published medical research remains a challenge. Two classic yardsticks are the citation count (the number of times a given paper is cited by others) and the impact factor of the journal that published the paper (which reflects the average number of citations per article). However, the citation count can be assessed only several years after publication, and the impact factor of the journal that published the paper (which reflects the average number of citations per article) is the “hit parade” posted on the journal website. Whether this count predicts citations is unknown. I examined this issue prospectively in a cohort of papers published in the BMJ.

Methods and results

The study used articles published in volume 318 of the BMJ (1999) in sections titled Papers, General Practice, and Information in Practice. The hit counts (full text articles, HTML version) for the main body of each article within a week of publication were provided by a BMJ staff member because the “hit parade” posted on the journal website was found to be unreliable for 1999. I obtained the number of citations on 24 May 2004 from the ISI Web of Science, an internet service to which the local medical library has a subscription.

I also recorded for each paper the study design and the number of pages. Nine papers were excluded because they did not report research (but reported discussions of, for example, NHS management and statistics methods). The remaining 153 papers comprised 29 randomised trials,
The value of a research study is traditionally assessed through citation counts or by the impact factor of the journal that published the study. Citation counts can be obtained only years after publication, and the impact factor is not paper specific.

What this study adds

For a cohort of papers published in the BMJ in 1999, the hit count on the website in the week after online publication predicted the number of citations in subsequent years; the hit count is a potentially useful measure of the scientific value of a research paper.

11 systematic reviews, 41 prospective studies, 8 case-control studies, 41 cross sectional surveys, 6 qualitative studies, and 17 other designs (such as economic analyses or case reports).

The average hit count for the papers in the first week after publication was 685 (SD 410; 25th, 50th, and 75th centiles 437, 578, and 795 respectively; range 175 to 3181); the average number of citations in the five years after publication was 32.5 (SD 37.5; 25th, 50th, and 75th centiles 9.5, 22, and 42.5 respectively; range 0 to 291). Only one paper was never cited. The hit count was associated with the number of subsequent citations (Pearson correlation coefficient: 0.50, P < 0.001). The result was similar for logarithms of the counts (r = 0.54, P < 0.001) (figure). For every 100 additional hits, 4.4 additional citations (95% confidence interval 3.1 to 5.7) accrued over the five years.

The average hit count for randomised trials or systematic reviews was 832, for prospective or case-control studies, 41 cross sectional surveys, 6 qualitative studies, and 75th centiles 9.5, 22, and 42.5 respectively; range 0 to 291). Only one paper was never cited. The hit count was associated with the number of subsequent citations (Pearson correlation coefficient: 0.50, P < 0.001). The result was similar for logarithms of the counts (r = 0.54, P < 0.001) (figure). For every 100 additional hits, 4.4 additional citations (95% confidence interval 3.1 to 5.7) accrued over the five years.

The average hit count for randomised trials or systematic reviews was 832, for prospective or case-control studies was 747, and for cross sectional, qualitative, and other studies was 545 hits (P = 0.001). Longer papers attracted more hits than short papers (an extra 7.3 citations per page, P < 0.001) and study design (randomised trials and systematic reviews yielded 46.0 citations, prospective and case-control studies 38.9 citations, and other designs 19.3 citations (P = 0.001). When the hit count was included as predictor, however, the effect of study design became non-significant; only page length (an extra 7.3 citations per page, P < 0.001) and the hit counts (an extra 3.7 citations per 100 hits, P < 0.001) remained as independent predictors. These variables explained 33% of variance in citation counts.

Comment

Papers that attracted the most hits on the BMJ website in the first week after publication were subsequently cited more often than less frequently accessed papers. Thus early hits counts capture at least to some extent the qualities that eventually lead to citation in the scientific literature.

My hypothesis is that “scientific value” explains the association between hits and citations. Online readers judge the scientific value of an article from the title and the abstract, and if this assessment is favourable, they access the full paper. The paper’s scientific value also leads to citation by other researchers. This hypothesis is supported by the greater frequency of both hits and citations for papers that used the most scientifically rigorous study designs, such as randomised trials.

The number of early hits is a potentially useful measure of the scientific value of published medical research papers. Publication of hit counts by online journals should be encouraged.

Daniel Berhane from the BMJ provided valid hit counts for the journal’s website.

Competing interests: TVP is the sole contributor.

Funding: None.

Ethical approval: Not required.

3 Seglen PO. Why the impact factor of journals should not be used for evaluating research. BMJ 1997;314:498-502.

(Accepted 10 July 2004)

Corrections and clarifications

Daniel Berhane from the BMJ provided valid hit counts for the journal’s website.

Competing interests: TVP is the sole contributor.

Funding: None.

Ethical approval: Not required.

3 Seglen PO. Why the impact factor of journals should not be used for evaluating research. BMJ 1997;314:498-502.

(Accepted 10 July 2004)