Assessment of Publication Metrics: Google Scholar vs PubMed Central

Kathleen Covino, PhD,[1] Piyali Dhar Chowdhury, PhD,[2] Ji-May Jen, PharmD,[2] Caitlin E. Lentz, PharmD,[3] Michele Patton, CMPP,[1] Melanie Blanchard, CMPP,[1] and Gina Fusaro, PhD[2]

[1]Nucleus Global, Hamilton, NJ, USA; [2]Celgene Corporation, Summit, NJ, USA; [3]Rutgers Institute for Pharmaceutical Industry Fellowships, Ernest Mario School of Pharmacy, Rutgers University, NJ, USA

ABSTRACT

Objective: Publication metrics are integral for assessing the impact of a publication in the peer-reviewed literature. We sought to compare the utility of 2 commonly used search engines, Google Scholar (GS) and PubMed Central (PM), for retrieval of citations in the peer-reviewed literature.

Methods: The citation of 2 phase 3 trials in metastatic breast cancer (MBCPub1 and MBCPub2) published in highly cited journals (*J Clin Oncol* and *N Engl J Med,* respectively) was followed from the time of publication using the "cited by" tools in PM and GS for the specific citation. The recall rate (ie, total number of clinical articles/total number of citations available), uniqueness of returns, and types of citations returned were assessed.

Results: For MBCPub1, GS and PM retrieved 321 and 46 items, respectively, of which 38 items appeared in both databases. Sixty-seven GS items (21%) were non–peer reviewed publications, including patents, theses, and continuing education materials. Of the 254 peer-reviewed publications retrieved; GS had a higher number of clinical trial publications (n = 46 vs n = 9), leading to a higher recall rate (18% vs 4%); Similar results were observed for MBCpub2, with 1250 GS and 243 PM items retrieved.

Conclusion: GS retrieved a higher overall number of items and proportion of peer-reviewed clinical articles compared with PM. Thus, GS may be a better option for determining publication reach. The use of both databases may provide the most accurate publication metrics.

BACKGROUND

- Publication metrics are used to monitor and assess the impact of a publication program and ensure the effective and efficient communication of key data to target audiences. The collection of appropriate data allows for accurate assessment of the performance of a publication plan
- Two publicly accessible online bibliographic databases, PubMed Central and Google Scholar, can be used to identify and retrieve information on the citation of a publication; however, these databases differ with respect to the scope and nature of publications indexed
  - Google Scholar is a search engine that indexes a broad variety of publications from various sources, whereas PubMed Central is an archive that indexes full-text articles published in the biomedical and life sciences literature

OBJECTIVE

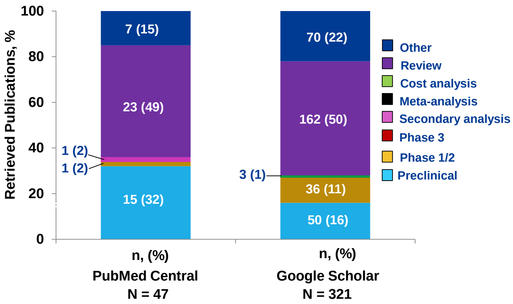
- To compare the utility of 2 commonly used search engines, Google Scholar and PubMed Central, for retrieval of citations in the peer-reviewed literature

METHODS

- We selected 2 publications of phase 3 clinical trials from our publication plan and comparable publications of phase 3 trials in metastatic breast cancer and advanced lung cancer published in high-impact journals (*J Clin Oncol* and *N Engl J Med*, respectively) for this analysis
- Citation of the publications in metastatic breast cancer (MBCPub1 and MBCPub2) or advanced lung cancer (LungPub1 and LungPub2) was followed using the "cited by" tools in PubMed Central and Google Scholar for the specific citation for either 2 or 5 years from the time of publication
- The number and types of citations identified were assessed using the following categories: preclinical, phase 1 or 2 trial, phase 3 trial, secondary analysis of a clinical study, cost analysis, meta-analysis, and review
  - An "other" category was used for theses, editorials/letters to the editor, supplements, web pages, case studies, book chapters, meeting summaries, newsletters, and patents that were retrieved by the search
- The recall rate (ie, calculated as total number of clinical trial articles [including phase 1, 2, 3 trials, secondary analyses of clinical study, and meta-analysis]/total number of citations available) and uniqueness of returns from each database were also assessed

RESULTS

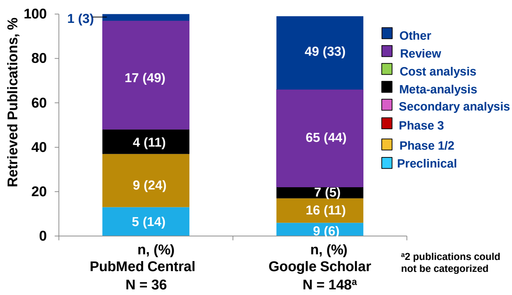
MBCPub 1 (*J Clin Oncol*) Citations, 2005 to 2010



- Google Scholar was more efficient than PubMed Central in recalling clinical trial publications. The recall rate for peer-reviewed clinical trial publications that cited MBCPub1 was higher for Google Scholar (n = 36; 11%) vs PubMed Central (n = 2; 4%)
- Although the number of review articles identified was higher in Google Scholar vs PubMed Central, the proportion of review articles retrieved by Google Scholar (n = 162; 50%) and PubMed Central (n = 23; 49%) was similar
- Of the 321 citing publications identified, 38 (12%) appeared in both databases
  - The majority of citations identified by PubMed Central also were retrieved by Google Scholar

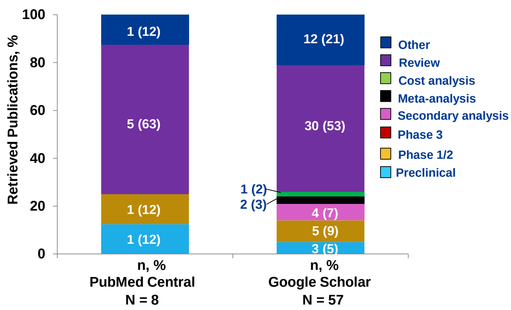
RESULTS (cont)

MBCPub 2 (*N Engl J Med*) Citations, 2007 to 2008



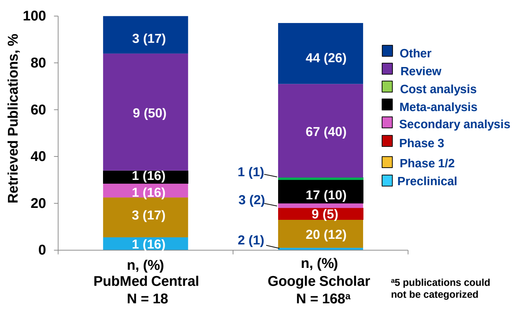
- The recall rate for peer-reviewed clinical trial publications that cited MBCPub2 was higher for Google Scholar (n = 23; 15%) vs PubMed Central (n = 13; 9%)
- Of the 148 citing publications identified, 36 (24%) appeared in both databases
- Citing publications in the "other" category identified by Google Scholar included supplements (n = 14), editorials/letters to editor (n = 9), websites (n = 6), book chapters (n = 3), and theses (n = 3)

LungPub 1 (*J Clin Oncol*) Citations, 2012 to 2014



- The recall rate for peer-reviewed clinical trial publications that cited LungPub1 was higher for Google Scholar (n = 11; 19%) vs PubMed Central (n = 1; 2%)
- Google Scholar retrieved a broader range and higher number of clinical trial publications, cost analyses, and meta-analyses
- Of the 57 citing publications identified, 8 (14%) appeared in both databases

LungPub 2 (*J Clin Oncol)* Citations, 2008 to 2009



- The recall rate for peer-reviewed clinical trial publications that cited LungPub2 was higher for Google Scholar (n = 49; 29%) vs PubMed Central (n = 5; 3%)
- Citing publications in the "other" category identified by Google Scholar included editorials/letters to the editor (n = 9), case reports (n = 8), workshop discussion/summary (n = 6), and book chapter, monograph, press release (n = 1 each)
- Of the 168 citing publications identified, 18 (11%) appeared in both databases

CONCLUSIONS

- Google Scholar retrieved a higher overall number of citing publications and proportion of peer-reviewed clinical articles compared with PubMed Central
  - Google Scholar was more likely to retrieve cost analyses, meta-analyses, and secondary analyses of phase 3 trials compared with PubMed Central
- Nonclinical trial publications retrieved by Google Scholar included editorials, case studies, workshop discussions, supplements, and book chapters
- The proportion of non-Medline-indexed publications retrieved by Google Scholar was low
- Based on our experience, Google Scholar appears to be a better option for determining publication reach. However, the use of both databases may provide the most accurate publication metrics

ACKNOWLEDGMENTS

The authors received printing support from MediTech Media, sponsored by Celgene Corporation. The authors are fully responsible for the content and editorial decisions for this poster.

Presented at the 10th Annual Meeting of ISMPP; Apr 7-9, 2014; Arlington, VA