Trends in biomedical informatics: most cited topics from recent years

Hyeon-Eui Kim, Xiaoqian Jiang, Jihoon Kim, Lucila Ohno-Machado

ABSTRACT
Biomedical informatics is a young, highly interdisciplinary field that is evolving quickly. It is important to know which published topics in generalist biomedical informatics journals elicit the most interest from the scientific community, and whether this interest changes over time, so that journals can better serve their readers. It is also important to understand whether free access to biomedical informatics articles impacts their citation rates in a significant way, so authors can make informed decisions about unlock fees, and journal owners and publishers understand the implications of open access. The topics and JAMIA articles from years 2009 and 2010 that have been most cited according to the Web of Science are described. To better understand the effects of free access in article dissemination, the number of citations per month after publication for articles published in 2009 versus 2010 was compared, since there was a significant change in free access to JAMIA articles between those years. Results suggest that there is a positive association between free access and citation rate for JAMIA articles.

INTRODUCTION
In fast moving fields such as biomedical informatics, it is important to know which topics attract the most interest of the scientific community. This information may be helpful to understand where gaps exist in the literature, and potentially guide investments in training and research. Through analysis of scientific literature, the trends in the biomedical informatics field may be revealed.

Access to information via the internet has changed the way scientific articles are written, and has made it easier for authors to reference other publications. PubMed\(^1\) has allowed anyone in the world who has access to the internet to freely procure articles of interest and obtain their abstracts. However, for articles published in journals with restricted access, obtaining the full article usually requires a subscription or pay-per-view fee.\(^1\)

While high impact journals are subscribed by a large number of institutions, journals that cover small specialties are restricted to fewer institutions. In order to promote growth and dissemination of relatively new fields such as biomedical informatics, it is important to understand whether free availability of articles significantly impacts their citation rates (which we use here as a proxy for dissemination of knowledge). We hypothesize that the unavailability of free access may contribute to a relatively low number of citations.

MATERIALS AND METHODS
We compared the keywords assigned in PubMed to 2009 and 2010 JAMIA articles to check whether there was a significant difference in frequency or citation rate per topic between articles published in each year. We performed the Kolmogorov–Smirnov to test the equality of two distributions of number of publications per highly cited topic in years 2009 and 2010.

We retrospectively evaluated the impact on Web of Science\(^2\) citations to JAMIA articles in those years, adjusting for the number of months a particular article was available for citation. For example, an article published in January 2009 has more chances of being cited than an article published in November 2010, so comparing total citation counts for these two articles requires normalization. We adjusted the number of Web of Science citations for all 2009 and 2010 articles according to the number of months that passed since publication date until November 1, 2011, the baseline date used for collection of the total number of citations.

Thus, the range of ‘months-since-publication’ covered 12 months (for articles published in November 2010) to 34 months (for articles published in January 2009), and was represented in two-month intervals (eg, 12, 14, 16, ..., 34 months), since JAMIA is a bimonthly publication. We used a unit of measure that reflects the number of citations per month (CPM)-since-publication. For example, an article that received 34 citations since publication in January 1, 2009 accounted for 1 CPM (34 citations/34 months), while an article that received 34 citations since its publication in November 1, 2010 accounted for 2.83 CPM (34 citations/12 months). In this example, the November 2010 article was considered more highly cited (higher CPM) than the January 2009 article, even though they had been cited the same number of times. A similar calculation was done for topics, defined by Medical Subject Headings (MeSH) terms. We also simulated what the citations to 2010 articles would be if they were available for 6 fewer months (specifically, we subtracted 6 months from the months since publication for the 2010 articles to accommodate the addition of 6 months to the embargo period).

We calculated the median and the range of CPM for articles published in 2009 and 2010, excluding 14 articles (1 in 2009 and 13 in 2010) that were offered free access since the day of their publication, and we performed the Wilcoxon test to evaluate the difference of CPM between the two years.
RESULTS

Figure 1A shows the number of articles on topics that were among the most highly cited. Not surprisingly, *Electronic Medical Records* is a frequent topic in both years, followed by *Methods, Information Storage and Retrieval, Natural Language Processing, and Organization and Administration*. There was no statistically significant difference in the distribution of topics between 2009 and 2010 (p=0.708). Hence, the confounding effect of topic on CPM was expected to be minimal.

The four most frequently cited topics, as judged by MeSH terms, were *Natural Language Processing, Medical Record Systems, Information Storage and Retrieval, and User-Computer Interface*. The most cited articles in both years within these categories were references 3-27 for *Medical Record Systems*, references 28-37 for *Information Storage and Retrieval*, references 38-42 for *User-Computer Interface* and references 43 for *Natural Language Processing*. The most cited articles were Reference 50 in 2009 and Reference 47 for 2010. *JAMIA*, *JAMA*, the *International Journal of Medical Informatics, Archives of Internal Medicine, Health Affairs*, and the *New England Journal of Medicine*.

The boxplot in figure 1B displays the median and ranges for the CPM in 2009, 2010, and 2010-simulated. The medians were 0.125, 0.071, 0.125 for 2009, 2010, and 2010-simulated, respectively. The ranges were 0.93, 0.86, and 2.13 for 2009, 2010, and 2010-simulated, respectively. There was 43% decrease in the median CPM from 2009 to 2010, and the difference was significant (p=0.00064) as calculated by the Wilcoxon test. There was no difference in the median from 2009 and 2010-simulated. The CPM for the highest-cited topics decreased between 2009 and 2010, with two exceptions: *Natural Language Processing and Methods* (see figure 1C).

One important factor that changed from 2009 to 2010 was *JAMIA*’s free access policy. The embargo period for all *JAMIA* articles increased from 6 months in 2009 to 12 months in 2010. We hypothesized that the decrease in CPM from 2009 to 2010 could be associated with this change in free access policy. Figure 2 shows the effect of this decrease of availability in terms of CPM. Figure 2A shows the CPM distribution for 2009 articles, while Figure 2B shows the distribution for 2010 articles. The top 20 articles, which correspond to nearly 20% of the total citations were compared, since a few, top articles dominate the citations and reputation of a journal according to Seglen.51 Among the 20 articles, 18 (90%) showed a decrease in CPM from 2009 to 2010 when the CPM difference was calculated in the same ranker in the two groups. To see the change in the trend in the top articles, a biexponential model, widely used in pharmacokinetics, was fit.52 53 We found that the fitted trend curve for 2010 decays much more rapidly than the fitted trend curve for 2009 (see figure 2D), which confirms the differences in medians depicted in figure 1B.

The discrepancy between 2009 and 2010 cannot be easily explained: the number of published papers in each topic was approximately the same (as shown in figure 1). Furthermore, the
same editorial team was in charge of article selection and hence we can assume that the article quality was the same. However, an important difference between 2009 and 2010 was the increase in the embargo period by 6 months. The distribution of CPM in 2010-simulated is much closer to the 2009 results, suggesting that the decrease in CPM in 2010 could potentially be explained by the 6-month increase in the embargo period.

**DISCUSSION**

The potential impact of free access on citations to biomedical informatics articles has seldom been studied. Although the scientific literature contains examples in which free access was shown to increase or, in some cases, decrease the number of citations to particular journals or to entire scientific disciplines, this has not been quantified for biomedical informatics articles.

Our analysis of citations has several limitations. First, the publisher changed from 2009 to 2010, and advertisement practices could have affected the dissemination of the articles and consequent citation rates. Second, there could be differences in article quality or topic popularity that we were not able to capture. Third, although the results are statistically significant, it is possible that the association does not depict a causal relationship. Fourth, most citation analyses consider periods of several years, which was not relevant in our case. Finally, the analysis is limited to *JAMIA* articles, and hence must not be generalized to other journals or to the field of biomedical informatics in general. Our goal was to understand the impact of free access on citation rates of articles recently published in

---

**Figure 2** Citation rates for 2009 and 2010 articles indexed by descending order of citations-per-month (CPM). (A) CPM for articles published in 2009. (B) CPM for articles published in 2010. (C) CPM for a simulated scenario in which we assumed that 2010 articles were available six fewer months, representing the increase in the embargo period until free access was granted. (D) Decay plot of CPM for 2009, 2010, and simulated 2010 scenario, fitted by a bi-exponential method.
CONCLUSION

JAMIA is an important publication venue for scientific articles reporting on innovative informatics research and applications. In the past three years, the journal has experienced significant changes in the length of its embargo period, and corresponding fluctuations in the number of citations. Currently, it is possible for JAMIA authors to unlock articles for immediate free access by paying a fee that is equivalent to the one charged by open-access journals, but few authors exercise this option. The reported data and analyses are far from exhaustive, but reveal potential trends that influence the role of the journal as a vehicle for wide dissemination for informatics innovations outside our own field. This information could be helpful in guiding policy decisions by journal owners and publishers.

Acknowledgments

The authors thank Dr Michele Day and Mr Daniel Capurro for insightful suggestions.

Funding

The authors are funded in part by US4 HL108460.

Competing interests None.

Provenance and peer review Commissioned; internally peer reviewed.

Data sharing statement Data are available as an online supplement at http://www.jamia.org.

REFERENCES


