Using Bibliometric Big Data to Analyze Faculty Research Productivity in Health Policy and Management

Christopher A. Harle, PhD [corresponding author]

Joshua R. Vest, PhD, MPH

Nir Menachemi, PhD, MPH

Department of Health Policy and Management,

Richard M. Fairbanks School of Public Health, Indiana University

1050 Wishard Blvd.

Indianapolis IN 46202-2872

PHONE: 317-274-5396

Keywords: big data, faculty productivity, bibliometrics, citation analysis

---

This is the author's manuscript of the article published in final edited form as:

Harle, C. A., Vest, J. R., & Menachemi, N. (2016). Using Bibliometric Big Data to Analyze Faculty Research Productivity in Health Policy and Management. The Journal of Health Administration Education; Arlington, 33(2), 285–293.

ABSTRACT

Bibliometric big data and social media tools provide new opportunities to aggregate and analyze researchers' scholarly impact. The purpose of the current paper is to describe the process and results we obtained after aggregating a list of public Google Scholar profiles representing researchers in Health Policy and Management or closely-related disciplines. We extracted publication and citation data on 191 researchers who are affiliated with health administration programs in the U.S. With these data, we created a publicly available listing of faculty that includes each person's name, affiliation, year of first citation, total citations, h-index and i-10 index. The median of total citations per individual faculty member was 700, while the maximum was 46,363. The median h-index was 13, while the maximum was 91. We plan to update these statistics and add new faculty to our public listing as new Google Scholar profiles are created by faculty members in the field. This listing provides a resource for students and faculty in our discipline to easily compare productivity and publication records of faculty members in their own and other departments. Similarly, this listing provides a resource for faculty, including department chairs and deans, who desire discipline-specific context for promotion and tenure processes.

INTRODUCTION

Objectively and consistently measuring individuals' research productivity is an important part of academic processes including hiring decisions, annual reviews, grant funding decisions, and tenure and promotion assessments. Despite many potential problems and caveats, the publication is a widely available and recognized unit of productivity (Petersen, Wang, & Stanley, 2010; Wilhite & Fong, 2012). Moreover, in an era of growing public scrutiny of academic effort and

activities and increased competition for funding resources, a listing of a researcher's publications provides one piece of quantitative evidence that we can easily observe.

Bibliometric analyses range from the rather crude to the complex. On one end of the spectrum is the raw count of published articles in a given timeframe (e.g., annually). Beyond counts, quality is often assessed by journal impact factor (Garfield, 2006) or journal reputation within the discipline (Brooks, Walker, & Szorady, 1991; Menachemi, Hogan, & DelliFraine, 2015). Other researcher-level metrics that are commonly used include a researcher's total number of citations, median number of citations, and the h-index (Hirsch, 2005). The h-index is defined as an individual researchers' number of publications "h" that have at least "h" citations. The h-index lends itself to being calculated over a person's career or for specific time periods (e.g., the last 5 years) and thus can facilitate assessments of one's performance over time. While h-index and other metrics each have their strengths and weaknesses (Bornmann & Daniel, 2007; Bornmann, Mutz, & Daniel, 2008; Hirsch, 2005, 2007), the increasing availability of public data and analytic tools make it easier to compute these metrics for many researchers, benchmark performance, and thus enable more objective and consistent comparisons within and across disciplines.

The purpose of the current paper is to describe the process and results we obtained after aggregating a list of Google Scholar profiles representing researchers in the Health Policy and Management discipline. We based our approach on a similar list created by computer science researchers for faculty in the area of biomedical informatics and other computer science-related fields (Lin & McCoy, 2016). We believe that the list will be useful to stakeholders in our

discipline in various ways. For example, researchers can use the list to help benchmark their academic productivity relative to their peers. Furthermore, prospective doctoral students or faculty candidates can use the list to compare productivity and publication records of faculty members in the departments to which they are applying. Similarly, department chairs, deans, and external letter writers can use the list to provide a discipline-specific and quantitative productivity assessment of faculty members who are being considered for tenure and/or promotion. Such assessments provide a contextualized perspective that can complement other quantitative as well as qualitative assessment tools. Finally, In addition to its utility for stakeholders in our discipline, this project is consistent with trends in health management and public health education and practice in that it represents an application involving novel "analytics," "big data", and social media technologies (Ellaway, Pusic, Galbraith, & Cameron, 2014; Fihn et al., 2014; Murdoch & Detsky, 2013).

METHODS

We used Web technologies to extract and display publication and citation data for 191 researchers health policy and management researchers from Google Scholar (scholar.google.com/). Google Scholar regularly aggregates, organizes, and makes publicly available bibliometric data from many publishers. Furthermore, individuals are able to create a public Google Scholar Profile that aggregates and displays a list of the articles they have authored, citation counts for each article, and bibliometric measures that summarize their body of work. These measures include total citations and h-index over their career and over the past five years. The Google Scholar profile for Albert Einstein, for example, shows a total of 94,659 citations and an h-index of 106 (Figure 1). In other words, Albert Einstein has 106 publications

that have each been cited at least 106 times.

We created initial list of potential faculty based on a previously-collected list of over 1,300 researchers who were invited to participate in the 2015 survey of Health Administration faculty survey (Menachemi, 2016). Next, we manually searched for each person on the list within Google Scholar and extracted the Web URLs for the 227 researchers with active Scholar profiles. Finally, we reviewed the Google Scholar profiles of these individuals and removed 38 whose primary research interests and publications we judged to be outside of health services, policy or management/administration. After creating the final list of 191 faculty and their Google Scholar profile pages, we adapted an existing, open-source utility that extracts citation statistics from public Google Scholar researcher profiles (Lin & McCoy, 2016). After extracting publication and citation data for each of these faculty, we created a public website that lists each person's name, affiliation, year of first citation, total citations, h-index and i-10 index (http://pages.iu.edu/~charle/). The i-10 index is the number of publications that have at least 10 citations. The listing also contains total citations, h-index, and i-10 index divided by the number of years since the researchers' first citation. The list is sortable by each of these metrics and searchable by name.

RESULTS

We extracted citation statistics for all 191 researchers on February 9, 2016 (Table 1). In this sample, total citations ranged from 1 to 46,363. The median of total citations was 700. The median researcher's total list of publications are cited on average 73 times per year. The most productive researcher, in terms of yearly citations, is cited 2,208 times per year. In our sample,

H-index ranged from 1 to 91, while the median h-index was 13. The overall distribution of h-index is right skewed with a peak around 9 and a tail that includes four relatively high h-indices of 60, 78, 79, and 91 (Figure 2).

DISCUSSION

Our use of publicly available big data provides a useful resource for understanding individual level research productivity in the Health Policy and Management discipline. We plan to regularly update the metrics and add new interested researchers to our bibliometric report, so that the website will be a timely resource for our discipline.

Our analyses are clearly applicable to researchers across the career span. For example, prospective doctoral students may be interested in comparing publications and impact of faculty at the programs to which they are applying. Indeed, counts and measures of citations and publications have been used to classify and rank programs in multiple disciplines (Kalaitzidakis, Stengos, & Mamuneas, 2003; Lazaridis, 2009). At the same time, such a listing as ours can serve as an exploratory tool for doctoral students. Doctoral students can visit the listing to help identify the leading scholars in the field, find relevant targets for publications, and gain an understanding of faculty productivity in various departments. For junior faculty members, these metrics can be a source to help set publication goals and benchmark their academic productivity relative to their peers. Similarly, department chairs, deans, and external letter writers can use the list to provide a discipline-specific and quantitative productivity assessment of faculty members who are being considered for tenure and/or promotion or during the hiring process. Such assessments provide a contextualized perspective that can complement other quantitative as well as qualitative

assessment tools.

It is important to note that our analysis is not a definitive description of scholarly activity within our discipline. In addition to the limitations inherent in measuring scholarly actively solely through publications, our sample of researchers may not be representative of all Health Policy and Management researchers. We derived the sample by starting with a list of faculty who are either affiliated with a health administration program recognized by the Association of University Programs in Health Administration (AUPHA) or who are members of the Academy of Management's (AOM) Health Care Management Division. While this list is extensive, it is probably not comprehensive. For example, some faculty members who identify as health policy or health management scholars may not affiliate with AOM and may reside in departments that do not have health administration programs. Second, many of the faculty members on our original list did not have public Google Scholar profiles and therefore could not be included. While it is unclear if researchers with certain individual characteristics or productivity levels are more or less likely to have a Google Scholar profile, this selection may also bias the representativeness of our sample. In the future, we expect to regularly update this list by adding new researchers and adding recent publication information. As the number of researchers in the listing grows, we expect the tool will become more comprehensive, more representative, and thus even more valuable. We encourage individuals who wish to be included on the list to create Google Scholar profiles and send the authors their profile URLs.

Another limitation of our Google Scholar analysis is that citation and h-index calculations for the same individual may vary between Google Scholar and other sources, such as Scopus or Web of

Science (Meho & Rogers, 2008; Meho & Yang, 2007). In particular, Google Scholar may include a broader set of publications, including more books, book chapters, dissertations, theses, reports, and conference presentations (Meho & Rogers, 2008). Also, we do not believe that the h-index, impact factors, publication counts, or any other single bibliometric measure provides a complete picture of faculty research productivity. Instead, these metrics each provide one (easily) quantifiable view of productivity. Still, the h-index has advantages over individual metrics, such as total citations or total publications. H-index is not inflated when a researchers has a small number of highly cited publications or when a researcher has a large number publications that are all infrequently cited (Bornmann & Daniel, 2007). However, researchers should cautiously compare h-indices between scholars who differ in how long they have been actively publishing. Clearly, a faculty member with a 25-year track publishing record is likely to have a higher h-index than even the most promising junior researcher. Researchers should also cautiously compare h-indices across fields or even sub-fields (Hirsch 2005). Given the interdisciplinary nature of Health Policy and Management, researchers within the field may span sub-fields that have different norms with respect to collaboration and authorship. Thus, it is possible that a researcher's h-index could be "earned" primarily through co-authorship as opposed to by regularly being the lead author on publications.

In conclusion, we created the first summary of Health Policy and Management faculty research productivity as measured using publicly available citation data. The listing exists as a public webpage, which we expect to update with new researcher profiles, publications, and citation data on a regular basis. The listing can be used to rank and otherwise analyze faculty in terms of h-index and other productivity metrics. We believe such analysis will provide valuable insights for

disciplinary researchers ranging from doctoral students to senior faculty.

Table 1. Summary citation statistics for researchers in Health Policy and Management

	Year of first citation	Total citations	Citations per year	H-index	i10-index
Mean (SD)	2001 (7.6)	2,513 (5,374.3)	161 (264.4)	17.7 (14.8)	31.3 (41.7)
Minimum	1977	1	1	1	0
25th %	1997	288.5	31	8	7
Median	2003	700	73	13	16
75th %	2008	2709	178	24	40
Maximum	2015	46363	2208	91	297

Note: Statistics based on data extracted from 191 public Google Scholar profiles on February 9, 2016.

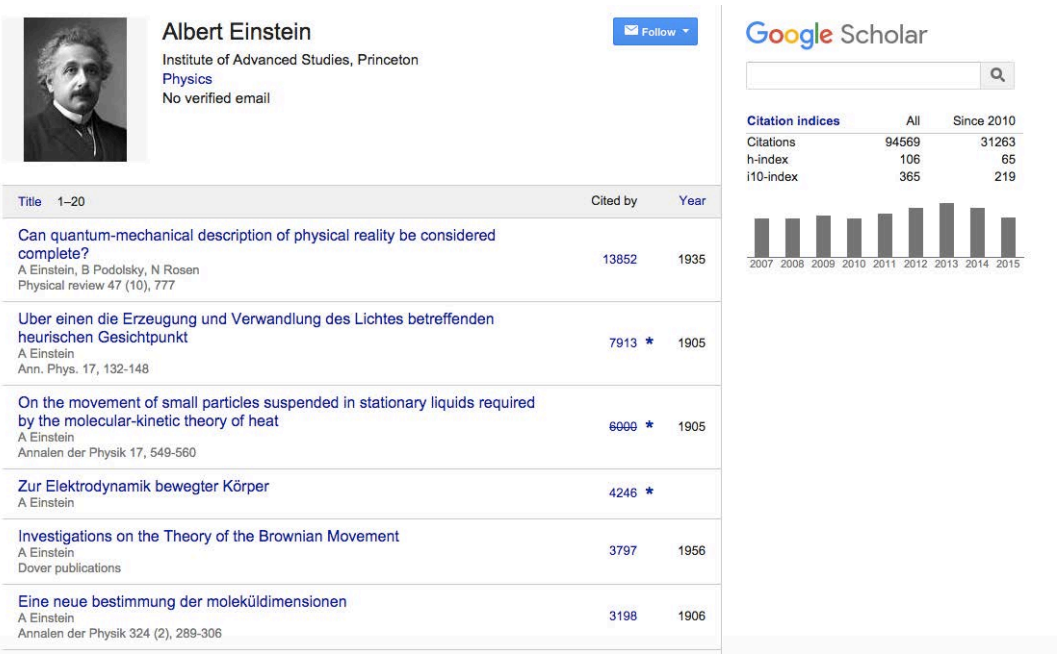

Figure 1. Example Google Scholar profile with citation statistics for Albert Einstein

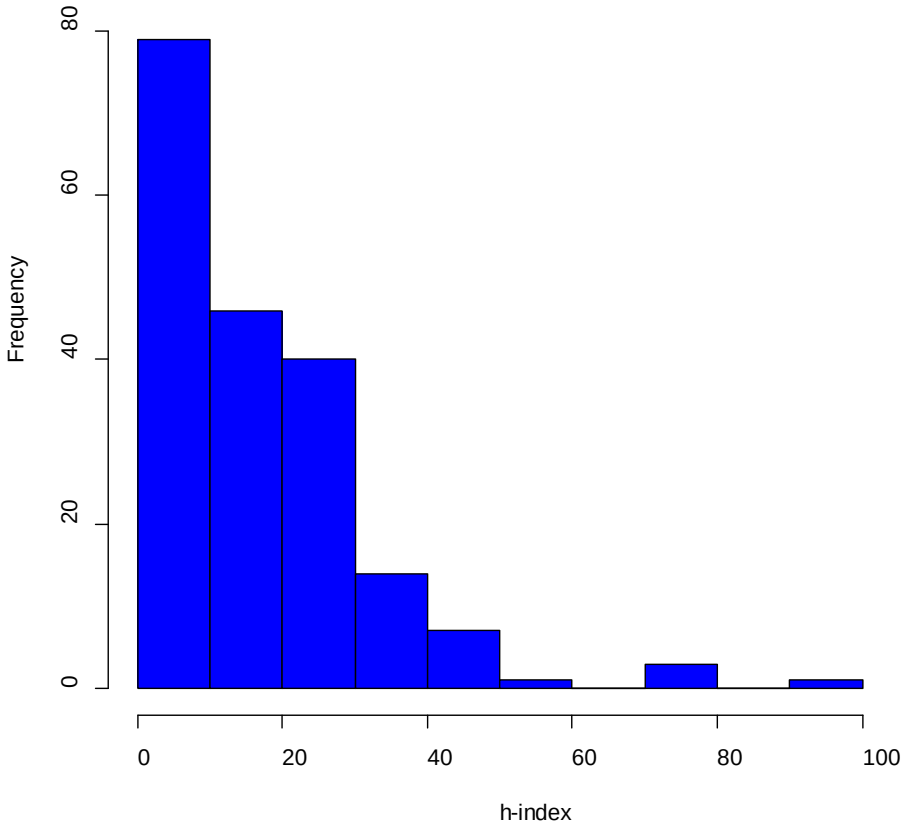

Figure 2. Distribution of h-index among 191 Health Policy and Management Researchers

REFERENCES

Bornmann, L., & Daniel, H.-D. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology, 58*(9), 1381-1385. doi: 10.1002/asi.20609

Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology, 59*(5), 830-837. doi: 10.1002/asi.20806

Brooks, C.H., Walker, L.R., & Szorady, R. (1991). Rating journals in health care administration: The perceptions of program chairpersons. *Medical Care, 29*(8), 755-765.

Ellaway, R.H., Pusic, M.V., Galbraith, R.M., & Cameron, T. (2014). Developing the role of big data and analytics in health professional education. *Medical Teacher, 36*(3), 216-222. doi: 10.3109/0142159X.2014.874553

Fihn, S.D., Francis, J., Clancy, C., Nielson, C., Nelson, K., Rumsfeld, J., . . . Graham, G.L. (2014). Insights from advanced analytics at the veterans health administration. *Health Affairs, 33*(7), 1203-1211.

Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA, 295*(1), 90-93. doi: 10.1001/jama.295.1.90

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569-16572.

Hirsch, J.E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences, 104*(49), 19193-19198.

Lin, J., & McCoy, A. (2016). Citations statistics of biomedical informatics researchers. Retrieved 9 Feb, 2016, from http://allisonbmccoy.github.io/scholar-scraper/index-bmi.html

Meho, L.I., & Rogers, Y. (2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of scopus and web of science. *Journal of the American Society for Information Science and Technology, 59*(11), 1711-1726. doi: 10.1002/asi.20874

Meho, L.I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of lis faculty: Web of science versus scopus and google scholar. *Journal of the American Society for Information Science and Technology, 58*(13), 2105-2125. doi: 10.1002/asi.20677

Menachemi, N. (2016). Research note: Overview of the data used in the current issue of JHAE and tests for response bias. Working paper.

Menachemi, N., Hogan, T.H., & DelliFraine, J.L. (2015). Journal rankings by health management faculty members: Are there differences by rank, leadership status, or area of expertise? *Journal of Healthcare Management, 60*(1), 17-28.

Murdoch, T.B., & Detsky, A.S. (2013). The inevitable application of big data to health care. *JAMA, 309*(13), 1351-1352. doi: 10.1001/jama.2013.393

Kalaitzidakis, P., Stengos, T., & Mamuneas, T.P. (2003). Rankings of academic journals and institutions in economics. *Journal of the European Economic Association, 1*(6), 1346-1366.

Lazaridis, T. (2009). Ranking university departments using the mean h-index. *Scientometrics, 82*(2), 211-216. doi: 10.1007/s11192-009-0048-4

Petersen, A.M., Wang, F., & Stanley, H.E. (2010). Methods for measuring the citations and productivity of scientists across time and discipline. *Physical Review E, 81*(3), 036114.

Wilhite, A.W., & Fong, E.A. (2012). Coercive citation in academic publishing. *Science, 335*(6068), 542-543.