Journal of Biomedical Informatics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Associating co-authorship patterns with publications in high-impact journals

Michael E. Bales [a,b], Daniel C. Dine [b], Jacqueline A. Merrill [a,c], Stephen B. Johnson [d], Suzanne Bakken [a,b,c], Chunhua Weng [a,b,*]

[a] Department of Biomedical Informatics, Columbia University, New York City, United States
[b] The Irving Institute for Clinical and Translational Research, Columbia University, New York City, United States
[c] School of Nursing, Columbia University, New York City, United States
[d] Weill Cornell Medical College, New York City, United States

ARTICLE INFO

Article history:
Received 2 December 2013
Accepted 11 July 2014
Available online xxxx

Keywords:
Co-authorship patterns
Social networking
MEDLINE
Journal impact factor
Evaluation

ABSTRACT

Objectives: To develop a method for investigating co-authorship patterns and author team characteristics associated with the publications in high-impact journals through the integration of public MEDLINE data and institutional scientific profile data.

Methods: For all current researchers at Columbia University Medical Center, we extracted their publications from MEDLINE authored between years 2007 and 2011 and associated journal impact factors, along with author academic ranks and departmental affiliations obtained from Columbia University Scientific Profiles (CUSP). Chi-square tests were performed on co-authorship patterns, with Bonferroni correction for multiple comparisons, to identify team composition characteristics associated with publication impact factors. We also developed co-authorship networks for the 25 most prolific departments between years 2002 and 2011 and counted the internal and external authors, inter-connectivity, and centrality of each department.

Results: Papers with at least one author from a basic science department are significantly more likely to appear in high-impact journals than papers authored by those from clinical departments alone. Inclusion of at least one professor on the author list is strongly associated with publication in high-impact journals, as is inclusion of at least one research scientist. Departmental and disciplinary differences in the ratios of within- to outside-department collaboration and overall network cohesion are also observed.

Conclusions: Enrichment of co-authorship patterns with author scientific profiles helps uncover associations between author team characteristics and appearance in high-impact journals. These results may offer implications for mentoring junior biomedical researchers to publish on high-impact journals, as well as for evaluating academic progress across disciplines in modern academic medical centers.

1. Introduction

Biomedical research is becoming increasingly interdisciplinary [1]. Numerous organizational factors have been recognized as barriers or facilitators of interdisciplinary research [2]. Although there are significant challenges in projects spanning multiple departments or disciplines [3], interdisciplinary research has been shown to be important for accelerating innovation [4].

A variety of analytical approaches, such as social–ecological models, systems thinking and complexity theories, social-determinants paradigms, and hierarchical analytic frameworks [5], have been employed to understand patterns of scientific collaboration. A prior bibliometric study has shown differences in co-authorship patterns across disciplines [6]. However, factors associated with the differences in scientific productivity have not been systematically quantified.

Given the central importance of scholarly publications and team-based scientific work, in this study we sought to understand scientific collaborations in biomedical research by investigating co-authorship patterns. Specifically, we sought to identify associations between co-authorship patterns and the impact factors of the journals of the publications. We leveraged the open-access Columbia University Scientific Profiles (CUSP) (http://irvinginstitute.columbia.edu/cusp) to obtain information about published researchers at our institution. Using CUSP, we enriched publication

* Corresponding author at: Department of Biomedical Informatics, Columbia University, New York City, United States.
E-mail address: cw2384@columbia.edu (C. Weng).

http://dx.doi.org/10.1016/j.jbi.2014.07.015
1532-0464/

data with institution-internal human resources data, including author academic rank and departmental affiliation. We employed two methodological approaches: analysis of authorship patterns and co-authorship networks. We then compared departments with respect to the ratio of within- to outside-department collaboration, as well as the overall levels of structural integration, all within our institution.

2. Materials and methods

2.1. Data sources and sample selection

Data were retrieved from our institution's research networking system, CUSP. CUSP was funded by Columbia University's Clinical and Translational Science Award (CTSA) to facilitate research networking and to help researchers identify experts and potential collaborators at CUMC. CUSP includes grants from institutional financial databases and publications from MEDLINE, along with job title, highest degree completed, and departmental affiliation from institutional human resource data. A core feature of CUSP is ReCiter [8], a method developed by the Columbia University CTSA for author name disambiguation for publications in scholarly databases. Researcher profiling systems often require investigators to populate their own publications manually. ReCiter keeps publications up to date by populating author publication lists automatically in CUSP through monthly feeds from MEDLINE. CUSP is interoperable with the open-source semantic web application VIVO, which enables the discovery of researchers across institutions [7].

When determining a time frame appropriate for article selection we sought to include enough articles to provide sufficient statistical power to address our research questions, while also minimizing the effects of missing data in older years. As CUSP only provides a snapshot of researchers currently employed at the university, historical data on academic rank and departmental affiliation were not available for current authors and no data is available for those who have left the university. Since personnel fluctuation is frequent in our university, it is appropriate to use a time period, e.g., a 5-year time period, that is shorter than our standard promotion time window (i.e., 7–11 years) for the analysis of patterns of authorship based on academic rank and departmental affiliations (Section 2.2) so that we can assume such information is less likely to have substantially changed during the short time frame.

Moreover, due to a recent major upgrade of administrative systems that provides departmental affiliation and rank to CUSP at our institution, year 2011 provided the most complete data at the time of the analysis. Therefore, we retrieved 7997 MEDLINE articles from 2007 to 2011 that included at least one author in CUSP. From this list of articles we identified 182 journals with an impact factor record, in which 10 or more articles were published during the time period. From these 182 journals we identified 3996 articles involving 2001 unique authors for this analysis. In contrast, for the co-authorship network analysis (Section 2.3), as social connections among researchers take time to develop, we sought to ensure that sufficient data on social links would be included. We therefore selected a 10-year period, 2002–2011, corresponding to a data set with 13609 articles, 2893 unique authors, and 2072 journals, from which individual co-authorship networks were generated for each of the top 25 departments in publishing volume (i.e., the 25 most prolific departments).

2.2. Co-authorship impact analysis

The first goal of this research was to characterize associations of author academic rank and departmental affiliation with publication in high-impact journals. After preliminary descriptive analysis we formulated our research questions as (1) what are the typical co-authorship patterns with respect to five specific author team properties (i.e., total number of authors, mixing of academic rank, inclusion of senior researchers, inclusion of junior scientists, and inclusion of basic or clinician scientists); and (2) which co-authorship patterns are associated with publications in high-impact journals?

To assign each article to a distinct journal impact tier we first ranked the articles based on journal impact factor for the year 2012, as reported in the ISI Journal Citation Reports [9]. We then divided the journals into three tiers based on journal impact rank: (1) High: [5.704, 51.658, $n = 60$]; (2) Medium: [3.371, 5.635, $n = 61$]; and (3) Low: [0.871, 3.320, $n = 61$]. We further labeled each article with one of these three journal impact categories.

We extracted academic rank and departmental affiliation for all authors having profiles in the CUSP system. For the analyses involving academic rank we included only investigators with an academic rank of postdoc, research scientist, assistant professor, associate professor, or professor (in our author academic rank notation, the term *professor* is used to denote full professors). Authors for whom academic rank was unavailable (e.g., authors at other institutions and researchers no longer employed at our university) were excluded. Authors were labeled according to their primary department affiliation.

In this context, authorship patterns are based on academic rank and on departmental affiliation. Possible combinations based on academic rank might be *one professor and one assistant professor*, or *one associate professor and two postdocs*. Similarly, possible combinations based on department type might be *one researcher from a clinical department and one researcher from dental medicine*, or *one researcher from public health and two researchers from basic science departments*. We enumerated author patterns for each paper as follows. First, we enumerated distinctive combinations of co-authors based on academic rank irrespective of author order. For example, if one paper had a professor as its first author, an assistant professor as its second author, and another professor as its third author, its academic rank pattern was PPI, representing *two professors* (*P*) *and one assistant professor* (I). More example patterns are provided below: (1) IP = one *assistant professor* and one *professor*; (2) OP = one *associate professor* and one *professor*; (3) IIP = two *assistant professors* and one *professor*; and (4) DP = one postdoc and one *professor*.

Second, we enumerated combinations of co-authors based on department type. A paper was considered to belong to a department type if at least one author on the paper was from the department type; as such, some papers included multiple department types. We calculated the number of departments involved on each paper. In this research we used the term *department* to refer to major organizational entities at our university, including departments within the school of public health, as well as basic science, clinical, and mixed basic/clinical departments within the medical school, interdisciplinary research centers that were classified administratively as departments, and the schools of Nursing and of Dental Medicine, which were not divided into departments. The distinction of basic vs. clinical vs. hybrid only relates to School of Medicine departments at Columbia University Medical Center (CUMC), where only basic science departments have Ph.D. programs. Clinical departments perform clinical services and research but cannot offer the Ph.D. Hybrid departments have Ph.D. programs and offer clinical services.

For each specific author academic rank and author department type combination we calculated numbers of articles published in high, medium, and low-impact journals. We assigned each paper into one of two categories along five separate axes: high (five or more) vs. low (four or fewer) numbers of authors; mixing of academic rank vs. single academic rank; inclusion of at least one professor vs. non-inclusion of professors; inclusion of at least one

Please cite this article in press as: Bales ME et al. Associating co-authorship patterns with publications in high-impact journals. J Biomed Inform (2014), http://dx.doi.org/10.1016/j.jbi.2014.07.015