Published in final edited form as:

J Cancer Educ. 2018 June ; 33(3): 564–568. doi:10.1007/s13187-016-1120-4.

Automated Literature Searches for Longitudinal Tracking of Cancer Research Training Program Graduates

Luz A. Padilla, M.D., M.S.P.H.¹, Renee A. Desmond, D.V.M., Ph.D.², C. Michael Brooks, Ed.D.³, and John W. Waterbor, M.D., Dr.P.H.¹

¹Department of Epidemiology, School of Public Health, University of Alabama at Birmingham.

²Division of Preventive Medicine, Department of Medicine, University of Alabama at Birmingham.

³School of Health Professions, University of Alabama at Birmingham.

Abstract

Background—A key outcome measure of cancer research training programs is the number of cancer-related peer-reviewed publications after training. Because program graduates do not routinely report their publications, staff must periodically conduct electronic literature searches on each graduate. The purpose of this study is to compare findings of an innovative computer-based automated search program versus repeated manual literature searches to identify post-training peer-reviewed publications.

Methods—In late 2014 manual searches for publications by former R25 students identified 232 cancer-related articles published by 112 of 543 program graduates. In 2016 a research assistant was instructed in performing Scopus literature searches for comparison with individual PubMed searches on our 543 program graduates.

Results—Through 2014 Scopus found 304 cancer publications, 220 that had been retrieved manually plus an additional 84 papers. However, Scopus missed 12 publications found manually. Together both methods found 316 publications. The automated method found 96.2% of the 316 publications while individual searches found only 73.4%.

Conclusions—An automated search method such as using the Scopus database is a key tool for conducting comprehensive literature searches, but must be supplemented with periodic manual searches to find the initial publications of program graduates. A time-saving feature of Scopus is periodic automatic alerts of new publications. Although a training period is needed and initial costs can be high, an automated search method is worthwhile due to its high sensitivity and efficiency in the long term.

BACKGROUND

Providing quality research training programs to medical students and health professional graduate students in public health, dentistry, and nursing, can stimulate their interest in

Corresponding author: John W. Waterbor, M.D., Dr.P.H., Associate Professor, School of Public Health, University of Alabama at Birmingham, 1665 University Boulevard, Birmingham, AL 35294-0022, h2obor@uab.edu, 205-934-7146.

The authors have no conflicts of interest to report with regard to this study.

engaging in research or continuing along a research career development path [1, 2, 3]. Some students believe that the time and attention required to engage in research may diminish their chances of acquiring a residency or advanced training. Other students believe that research experience is useful, but only to build their resume's in order to increase their chances of selection for post-graduate training [4,5,6]. These attitudes and beliefs may obscure the benefits that research experience could provide for the career development of medical and graduate health professional students. Such students who participate in research training programs acquire a better understanding of research methods and results, allowing them to make better professional and clinical decisions as compared to their peers [7]. Research training also enhances critical thinking, improves the quality of training, provides new skills, and sensitizes students to health-related issues [8]. A recent study of the impact of intensive mentoring and structured training for early career clinical faculty demonstrated that the number of research grants and dollar amounts of the grant awards post-training was about three times greater than for a comparison cohort not similarly mentored and trained [9].

Research training programs have proven effective in increasing the number of research-oriented health professionals [10,11]. The National Institutes of Health provides funding for research training programs through its R25 program, available in 27 institutes including the National Cancer Institute (NCI). In 1988 UAB's Cancer Prevention and Control Training Program (CPCTP) for the support of a post-doctoral fellows, doctoral students, and a few summer students each year, was funded by NCI [12-14]. In 1999 UAB's Cancer Research Experiences for Students (CaRES) program, a spinoff of the CPCTP for summer students only, was funded by NCI through its R25E program for short-term research projects [15-17]. The CaRES program's primary goal is to introduce or re-introduce eligible students to cancer research and to motivate them to consider cancer research careers. To date, the CaRES program has awarded over 600 summer internships in cancer research for UAB medical students and graduate health professional students who are American citizens or permanent residents of the U.S. The proportions of CaRES students, according to academic program were 57% from medicine, 29% from public health, 7% from dentistry, and 7% from nursing and related programs.

A complete description of the CaRES program is found in our 2015 paper [17]. Each CaRES research project lasts from eight to twelve weeks, based on the demands of the project and the availability of the preceptor and student. Students complete a mid-summer evaluation in order to identify issues of concern that can be corrected before summer's end. Exit evaluations of students and preceptors measure satisfaction with their research experiences and with program administration. Short-term evaluation results over the past several years are presented on our program website, www.uab.edu/cares. Since its inception CaRES has met its short-term goals as documented by outstanding exit evaluations by students and preceptors [17].

In order to evaluate whether CaRES is meeting its long-term goal of motivating students to consider cancer research in their career plans, program directors must assess the number and quality of cancer-related publications, abstracts, and posters co-authored by program graduates; and assess the graduates' achievements in cancer research, patient care and in other professional roles related to the program's goals. The purpose of this paper is to

describe an automated search approach for identify post-training publications of program graduates, and to compare our findings with the results of periodic, repeated manual literature searches on each former student, in terms of 1) start-up time; 2) personnel effort; and 3) sensitivity for finding relevant publications.

METHODS

IRB

Tracking alumni of our university is not subject to approval by our Institutional Review Board, and therefore this study did not require IRB review and approval. No personal identifying information on alumni are presented in this report as only summary data are analyzed.

Database training

CaRES personnel met with one of our institution's medical librarians to learn about options for subscribing to an automated system of alerts to inform us when program graduates publish peer-reviewed papers. We learned that several such systems exist, and that these systems can be used to identify career-long publications of program participants. The efficacy of automated searches for career publications is the subject of this study. The librarian presented the strengths and limitations of Scopus [18, 19], PubMed [20], Web of Science [21], and Google Scholar [22]. We chose Scopus, to which our university subscribes, as the search database that best fit our needs. Launched in 2004 by Elsevier Publishing Company, Scopus is the largest database for abstracts and citations of peer-reviewed literature on technology, medicine, art, humanities and social sciences, extending back to the 1960's [20]. Our goal was to program the Scopus system to set publication alerts for the 543 CaRES graduates who participated in the program from 1999 to 2014. Were this approach effective, it would diminish or eliminate the need for periodic individual manual searches on each graduate. The medical librarian reviewed the mechanics and database platform of Scopus with our research assistant, and a few trial searches were conducted until the research assistant became comfortable with the Scopus process.

Following initial training we uploaded our list of contact information on all CaRES program graduates, so that we could match the names in our database to author names found on Scopus searches. This list was updated in late 2014 for a longitudinal tracking survey of 1999-2013 alumni. Contact information obtained from the alumni affairs offices of the university and Schools of Medicine and Public Health included name (and its variations), last known institution where a student or employee, last known physical address, phone numbers, email addresses, and residency locations (for medical students). We became aware of individuals who had changed their name (by virtue of marriage, divorce, or for other reasons) and may publish under a name other than their student name. Scopus accounts for author name variations and changes as long as all versions pertain to the same author identification number (ID). After copying the list of names to a Word document, we set "alerts" to notify us whenever individuals by these names publish a paper. Notifications were to be sent to the university email addresses that we selected. At our option, alerts were programmed to be sent monthly, as opposed to quarterly.

For the names in our CaRES database, “Author Search” on Scopus notified us of all publications to date, and any new publications that it finds as the months go by. Scopus is programmed to list all publications that were authored by someone having the same last and first name. Our knowledge of the area of study (cancer), institutional affiliation, and place of employment of our CaRES students, allow us to confirm, or fail to confirm, matches with the authors found by Scopus. Results are filtered by these components. Once a possible match is found, the author profile is reviewed by the program director and research assistant in order to identify and delete apparent matches that are incorrect. We deleted publications unrelated to cancer and publications from years prior to the CaRES summer training period. The list of publications found on Scopus for the period 1999-2014 were cross-referenced with the CaRES publication database from 2014.

Statistical analysis

There is no gold standard for identifying all relevant publications for each program participant. The primary reason for this is that each database search engine operates on a proprietary algorithm and it would be cost-prohibitive to test multiple search engines for our cohort. Therefore the “universe” of relevant publications for this study is defined as the sum of all publications identified by both methods. The sensitivity for each method is defined as the number of publications discovered by that method divided by the total number. The relevant exact 95% confidence interval can be computed for automated and for manual searches.

RESULTS

Start-up time

After two hours of training by a medical librarian our research assistant became facile in navigating the Scopus system. Inputting and checking the identifying information on 543 CaRES students took approximately ten hours, accounting for double-checking dubious spellings and contact information, and entering multiple versions of student names (e.g., with or without a middle initial).

Personnel

Our research assistant was extremely capable and did most of the hard work, with oversight by the CaRES program director who resolved problems and lent guidance as needed.

Quantitative results

Of 543 CaRES students we were able to set Scopus alerts for the 278 individuals (51%) who had published at least one paper. Alerts could not be set for the remaining 265 CaRES students who had not published to date. These remaining alumni must be manually searched until they publish their first paper and can be added to the Scopus system. Table 1 shows the number of cancer-related peer-reviewed publications by former CaRES students each year post-training, comparing manual search results with Scopus results. Scopus consistently found publications missed by the manual search method, but the manual method occasionally found publications missed by Scopus. Through 2014 Scopus found 304 cancer-related publications, 220 that had been found manually plus 84 more. However, Scopus

missed 12 publications that had been found by manual searches. By using both methods we found a grand total of 316 publications: 220 found by both methods plus 12 (5% more) found only manually plus 84 (38% more) found only by Scopus. Using the denominator of 316 total publications, the sensitivity of the Scopus search was significantly higher, 96.2% (93.5, 98.0) versus the manual search sensitivity of 73.4% (68.2, 78.2). The 12 papers missed by Scopus were published in a variety of journals, and most had PMCID numbers. It is unclear why these papers were missed by Scopus. Of the 84 publications missed by manual searches, many were published in journals of high quality, many of them surgery journals. Why this large number of papers was repeatedly missed on manual searches is unclear, and raises methodological concerns.

DISCUSSION

One aspect of evaluating long-term achievements of research training programs is the number of peer-reviewed publications by students who complete the program. Individual literature searches on dozens or hundreds of program graduates can be tedious and time-consuming, and must be repeated periodically. Automated search methods are now available through existing databases, such as Scopus. This approach may decrease the burden of gathering statistics for annual reports and training grant proposals. Once programmed, Scopus can provide immediate alerts of new publications of interest, without having to perform future, repeated searches on individual authors. Scopus is used not only in academic settings but by government agencies as well, and it is the largest resource for accessing abstracts and citations of peer-reviewed literature [23]. This is the primary reason why we chose Scopus as well as the fact that our institution has a subscription. Due to limitations of time and money, an active comparison of Scopus was not made against other options such as Web of Science [24], a pioneer in the field of literature searches. Such comparative studies may be useful for institutions contemplating subscribing to an automated search database.

Scopus has many positive attributes. For individuals who have published, Scopus can establish how many papers have ever been published, whether or not in an area related to the training program. Scopus accounts for variations in author names, although it is important to input all alternate names to ensure that searches are conducted for all possible name variations. Knowing a student's middle initial, institutional affiliation or place of employment can differentiate him or her among authors having identical or near-identical names, minimizing search time. This is most useful in searching short, common author names. Scopus may find many authors with matching first and last names, requiring additional information from the filters to confirm a match. An alert can be set for two or more authors in combination, for instance, a student and his or her preceptor. A previous list of publications is useful to identify past co-authors and thus, possible co-authors of new publications. After setting a Scopus automated alert, there is no future need for manual searches on that graduate. A program's student publication list can be updated as frequently as monthly, and Scopus has no limit on the number of alerts that can be set by a user.

Scopus has its limitations. An author alert can be set only for individuals who have published at least one paper. The manual search process must be repeated periodically for individuals who have never published, to determine when or whether that first publication

appears. Scopus may not be useful for research training programs whose students have few if any previous publications. Search results may vary depending on the ability and skills of the Scopus user. The author alert is typically linked to an Open Researcher and Contributor ID (ORCID) number, and if for any reason this number changes, a new search may be required. A subscription to Scopus or other automated search approaches can be costly. Prices are not publicized and vary according to the buyer and the size of the institution. Systems can cost thousands of dollars annually for large institutions. Because of the high cost it is unlikely that an institution has subscriptions to two or more databases, making comparisons of results difficult [24] unless there is multi-institutional collaboration. Software available in the marketplace is constantly changing and the best option for a program may change over time; the choice should be made with care.

CONCLUSIONS

Setting up automated literature searches may be essential for large cancer research training programs that must keep their program publication lists updated for purposes of annual reports and grant proposals. Automated literature searches can save time and increase efficiency as compared to individual searches on program participants, which must be updated periodically. Start-up time is considerable but is not repeated (other than for adding new author names) once alerts are set. Personnel efforts are minimal in the long-term, compared to repeated individual manual PubMed searches. Publication “catchment” or sensitivity can be much higher than for manual searches. In our study Scopus found 38% more publications than did our previous repeated manual searches, although Scopus missed a few publications found manually. Automated search methods can be supplemented with individual literature searches on program graduates who have never published, and should be done to find the occasional publication missed by the automated method.

ACKNOWLEDGEMENTS

The authors wish to thank Ms. Lee Vucovich of UAB’s Lister Hill Library of the Health Sciences and Ms. Judy Baker of the Department of Epidemiology in the UAB School of Public Health for their guidance on and assistance with this project. This research was supported in part by UAB’s Cancer Research Experiences for Students (CaRES) Program, 5R25CA076023, funded by the National Cancer Institute.

REFERENCES

1. Farooq Sheikh AS, Sheikh SA, Kaleem A, Waqas A. Factors contributing to lack of interest in research among medical students. *Adv Med Educ Pract.* 2013; 4:237–243. Published online Nov 7, 2013. PMID: PMC3826903. DOI: 10.2147/AMEP.S51536 [PubMed: 24235856]
2. Harsha Kumar H, Jayaram S, Kumar GS, Vinita J, Rohit S, Satish M, Shusruth K, Nitin A. Perception, practices towards research and predictors of research career among UG medical students from coastal south India: a cross-sectional study. *Indian J Community Med.* 2009; 34(4):306–9. [PubMed: 20165623]
3. Zier K, Wyatt C, Muller D. An innovative portfolio of research training programs for medical students. *Immunol Res.* 2012; 54(1-3):286–291. Review. PMID: 22418729. DOI: 10.1007/s12026-012-8310-x [PubMed: 22418729]
4. Jimmy R, Palatty PL, D’Silva P, Baliga MS, Singh A. Are medical students inclined to do research? *J Clin Diagn Res.* 2013; 7(12):2892–5. PMID:24551667. DOI: 10.7860/JCDR/2013/6698.3786 [PubMed: 24551667]

5. Burgoyne LN, O'Flynn S, Boylan GB. Undergraduate medical research: the student perspective. *Med Educ Online*. 2010; 10:15. PMID: 20844608. doi: 10.3402/meo.v15i0.5212
6. Nikkar-Esfahani A, Jamjoom AA, Fitzgerald JE. Extracurricular participation in research and audit by medical students: opportunities, obstacles, motivation and outcomes. *Med Teach*. 2012; 34(5):e317–24. DOI: 10.3109/0142159X.2012.670324 [PubMed: 22471919]
7. Committee C, Institute for International Medical Education. Global minimum essential requirements in medical education. *Med Teach*. 2002; 24(2):130–5. [PubMed: 12098431]
8. Naqvi HA. Students' research: tradition ahead of its time. *J Coll Physicians Surg Pak*. 2010; 20(10): 701–2. DOI: 10.2010/JCPSP.701702 [PubMed: 20943120]
9. Libby AM, Hosokawa PW, Fairclough DL, Prochazka AV, Jones PJ, Ginde AA. Grant success for early-career faculty in patient-oriented research: difference-in-differences evaluation of an interdisciplinary mentored research training program. *Academic Medicine*. Jun 21.2016 2016. doi: 10.1097.
10. Devi V, Abraham RR, Adiga A, Ramnarayan K, Kamath A. Fostering research skills in undergraduate medical students through Mentored Student Projects: Example from an Indian medical school. *Kathmandu University Medical Journal. Kathmandu Univ Med J*. Jul-Sep;2010 8(31):294–8. DOI: 10.3126/kumj.v8i3.6215
11. Hunskaar S, Breivik J, Siebke M, Tømmerås K, Figenschau K, Hansen JB. Evaluation of the medical student research programme in Norwegian medical schools. A survey of students and supervisors. *BMC Med Educ*. 2009; 9:43. [PubMed: 19602226]
12. Heimbürger DC, Waterbor J, Fish L, Brooks CM. Cancer prevention and control training program at the University of Alabama at Birmingham. *J Cancer Educ*. 2000; 15(2):69–72. [PubMed: 10879893]
13. Heimbürger DC, Waterbor JW, Fish L, Brooks CM. An interdisciplinary training program in nutrition sciences and cancer. *J Cancer Educ*. 2000; 15(3):130–3. [PubMed: 11019757]
14. Waterbor JW, Heimbürger DC, Fish L, Etten TJ, Brooks CM. An interdisciplinary cancer prevention and control training program in public health. *J Cancer Educ*. 2002; 17(2):85–91. PubMed [journal] PMID:12092859. [PubMed: 12092859]
15. Daniel CL, Brooks CM, Waterbor JW. Approaches for longitudinally tracking graduates of NCI-funded short-term cancer research training programs. *J Cancer Educ*. 2011; 26:58–63. [PubMed: 21210271]
16. Padilla LA, Venkatesh R, Daniel CL, Desmond RA, Brooks CM, Waterbor JW. An Evaluation Methodology for Longitudinal Studies of Short-Term Cancer Research Training Programs. *J Cancer Educ*. 2016; 31(1):84–92. NIHMSID: NIHMS644185 PubMed [journal] PMID:25412722, PMCID: PMC4441613. [PubMed: 25412722]
17. Desmond R, Padilla LA, Daniel CI, Prickett CT, Venkatesh R, Brooks M, Waterbor JW. Career outcomes of graduates of R25E short-term cancer research training programs. *J Cancer Educ*. Jan 22.2015 [Epub ahead of print].
18. Scopus. Journal Metrics: About Scopus. Elsevier; 2016. <http://www.journalmetrics.com/about-scopus.php> [Accessed 2 February 2016]
19. Elsevier. [Accessed 2 February 2016] Who uses Scopus: The A&I research discovery solution for academia, business and government. 2016. <https://www.elsevier.com/solutions/scopus/who-uses-scopus>
20. PubMed. [Accessed 2 February 2016] US National Library of Medicine National Institutes of Health. 2016. <http://www.ncbi.nlm.nih.gov/pubmed>
21. Web of Science. Thomson Reuters; 2016. Journal Citation Reports and Essential Science Indicators. https://apps.webofknowledge.com/select_databases.do?highlighted_tab=select_databases&product=UA&SID=3DPXjiCoH8IHmwVXk9&cacheurl=no&errorKey=errors.noProductSubscription#searchErrorMessage [Accessed 2 February 2016]
22. Google Scholar. [Accessed 2 February 2016] 2016. <https://scholar.google.com/>
23. Dess, H. [Accessed 15 February 2016] Database Reviews and Reports. Issues in Science and Technology Librarianship. 2016. <http://www.isrl.org/06-winter/databases4.html>

24. Fingerman, S. [Accessed 30 June 2016] Web of Science and Scopus: Current Features and Capabilities. *Issues in Science and Technology Librarianship*. 2006. <http://www.istl.org/06-fall/electronic2.html#3>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Papers found by Scopus Search, Manual PubMed Search, and by both methods.
*Image created and edited using tools from Microsoft Office

Table 1

Cancer related peer-reviewed publications by CaRES alumni after traineeship by year

Year	Total found by Manual and/or Scopus Search	Manual PubMed Searches	Scopus Search	Found only by manual PubMed searches	Found only by Scopus	Papers found by both sources
1999						
2000						
2001						
2002	2	2	2	-	-	2
2003	5	4	5	-	1	4
2004	6	5	6	-	1	5
2005	9	5	9	-	4	5
2006	15	9	15	-	6	9
2007	21	15	20	1	6	14
2008	15	10	15	-	5	10
2009	24	15	24	-	9	15
2010	21	15	21	-	6	15
2011	40	33	37	3	7	30
2012	43	35	43	-	8	35
2013	50	39	45	5	11	34
2014	65	45	62	3	20	42
Total	316	232	304	12	84	220