

EDUCATION
RESEARCH ARTICLE

Using publication data to evaluate a Clinical and Translational Science Award (CTSA) career development program: Early outcomes from KL2 scholars

Noelia Sayavedra^{1,2*}, Janice A. Hogle³ and D. Paul Moberg^{1,2}

¹ Institute for Clinical and Translational Research, University of Wisconsin, Madison, WI, USA

² Population Health Institute, University of Wisconsin, Madison, WI, USA

³ Hogle Slater Consulting, Crozet, VA, USA

Journal of Clinical and Translational Science (2017), 1, pp. 352–360 doi:10.1017/cts.2018.1

Introduction. This study uses KL2 scholars' publications to evaluate the types of research the KL2 program supports and to assess the initial productivity and impact of its scholars.

Methods. We illustrate the feasibility of 3 different approaches to bibliometrics, one viable method for determining the types of research a program or hub supports, and demonstrate how these data can be further combined with internal data records.

Results. Gender differences were observed in the types of research scholars undertake. Overall KL2 scholars are performing well, with their publications being cited more than the norm for National Institutes of Health publications. Favorable results were also observed in scholars' continued engagement in research.

Conclusion. This study illustrates that linking bibliometric data and data categorizing publications along the translational spectrum with a Clinical and Translational Science Award hub's internal data records is feasible and offers a number of innovative possibilities for the evaluation of a Clinical and Translational Science Award hub's programs and investigators.

Received 5 October 2017; Revised 30 November 2017; Accepted 13 December 2017

Key words: Career development, CTSA, translational research, bibliometrics, publications.

Introduction

The goal of National Institutes of Health's (NIH) Clinical and Translational Science Award (CTSA) program is to facilitate and accelerate the translation of discoveries from the bench to the bedside and into communities. In order to accomplish this, clinical and translational researchers must be supported and developed [1]. As a result, training and career development have been core components of the CTSA program since its launch in 2006. NIH and CTSA hubs all stress the importance of developing the next generation of clinical and

translational researchers. There has been a considerable investment in training and career development programs, as evidenced by time, funds, and the growing scope and number of such programs. However, there is a lack of evaluations of CTSA training and career development programs aimed at understanding the impact such programs have on research productivity and ultimately on career success.

The pilot evaluation project discussed in this paper focuses on one such career development program, specifically the KL2 program. The KL2 is a multi-year career development program that provides salary support and funds for protected training and research time for early-career scholars (M.D., M.D./Ph.D., or Ph.D.), allowing them to conduct research while being mentored by experienced investigators [2]. This article looks at the publication patterns, scholarly productivity, and research impact of the KL2 scholars' published work at one CTSA hub—the Institute for Clinical and Translational Research (ICTR) at the University of Wisconsin-Madison (UW)—using bibliometric data, data on research translational stage, and program-level data detailing scholar characteristics. Within the scope of this paper, we discuss the

* Address for correspondence: N. Sayavedra, Population Health Institute, University of Wisconsin-Madison, 610 Walnut Street, 911 WARF, Madison, WI 53726, USA.
(Email: nsayavedra@wisc.edu)

innovative methods we undertook for analyzing the data and the outcomes, as well as the challenges associated with this type of analysis.

Although complex models of career success have been put forward, such as the comprehensive career-success model for physician-scientists in Rubio *et al.* [3] crafted by the Research on Careers Workgroup at the University of Pittsburgh, the scope of this pilot study is more limited. We focus on aspects of extrinsic career success, specifically on concrete outcomes and markers such as publications [3]. Within the scope of this pilot study, we were also able to explore some personal contextual factors/determinants of career success, specifically demographic characteristics such as gender [3, 4].

Current Challenges

The challenges inherent in evaluating training and career development programs even at the single institutional level are daunting. These multiply when the focus is expanded across CTSA hubs which is likely a major reason for the lack of this type of evaluation.

The lack of nationally coordinated/consortium-level evaluative efforts and a dearth of standardized methods or shared definitions are reoccurring problems that have hindered such efforts. For instance, all CTSA hubs are required to report on publications by scholars who received hub funding for career development such as KL2 and TLI awards, as well as other publications that are the product of research that used CTSA resources and services. Yet despite having access to the necessary publication data (via required reporting) and the tools needed to undertake bibliometric analyses via their iCite tool, the NIH does not analyze and produce reports using bibliometrics. A consortium-level analysis of KL2 scholars would be valuable and could be used to provide insights into trends in workforce development and provide a useful benchmark that CTSA hubs could use to help evaluate their programs. However, the lack of a nationally coordinated plan that utilizes shared definitions and a standardized methodology means that, in general, CTSA hubs have been left to create their own tracking processes and determine what to include and measure in order to determine the impact of their programs [5].

Other instances where a lack of guidance and shared definitions can pose challenges include issues touched upon in the Institute of Medicine (IOM) report from the 2013 assessment of the CTSA program which recommends CTSA hubs should support research across the translational spectrum [6]. This recommendation, while important and well-intentioned, first requires that a CTSA hub have a means of determining exactly the types of research they are supporting before they can begin to weigh their relative success in supporting the full spectrum of translational research. The 2013 IOM report acknowledges that greater standardization is needed and that metrics need to go further than simply counting publications, yet the report did not offer specific guidance on what other metrics should be used [5, 6]. Recently attempts have been made to address some of the issues discussed above through the phased development and rollout of a Common Metrics Initiative which may ultimately provide more guidance on shared methods and definitions across the CTSA hubs.

Examination of a hub's publications and associated citation network, as well as the publications associated with its specific services and programs (such as the KL2 program), has the potential to contribute to a richer understanding of the quantity, impact and types of research a CTSA hub supports. However, moving beyond publication analysis to citation analysis is challenging. Citation analysis typically requires not only that hubs have access to citation databases, but also that publication and citation data be compiled or acquired in a format that allows for manipulation—a tall order [7].

An additional compounding factor is the issue of normalization, as indicators should be normalized in order to make comparisons between publications that are published in different fields, years, and journals.

However, normalized citation impact indicators, specifically field-normalized indicators, are not typically offered for free or in an easy-to-use exportable format by traditional providers such as Thomson Reuters and Elsevier, putting the burden of calculating such indicators on the user. Calculating normalized citation impact indicators can be both time-consuming and complex, depending on the ease of access to the additional data needed for the calculation and the complexity of calculation itself. Because of the challenges associated with conducting citation analysis, many turn to paid services offered by Thomson Reuters and Elsevier to gather and analyze their publication and citation data. However, such service providers interfaces, in the past, have tended to be inflexible. The inflexible nature of these interfaces can limit their usefulness when one desires to drill down into the data, focus on smaller subsets of data such as those associated with individual programs, or merge in additional data.

This pilot represents one CTSA hub's first efforts to contribute to an evaluation of a career development program by addressing many of the challenges mentioned above. We combined multiple data sources, and used KL2 scholars' publications as the basis of an evaluation that utilizes internal data on scholars' demographics, translational research category codes, and bibliometrics. Using this approach, we were able to take a preliminary look at the types of research the KL2 program supports and assess the initial productivity of its scholars and the impact of their published research.

Methods

This pilot study provides an overview of the early scholarly contributions/productivity of ICTR KL2 scholars at the UW-Madison using publication data. In doing so, we illustrate the feasibility of 3 different approaches to bibliometrics: utilizing paid services (specifically custom reports through providers such as Thomson Reuters or Elsevier), using the basic free downloadable reports or data file options provided by Thomson Reuters or Elsevier, or using the free services of newer options such as iCite. In addition, to illustrating the feasibility of conducting bibliometric analysis using 3 different data sources/approaches, we also illustrate how bibliometrics can be used to conduct analyses at both the publication level and author level. Furthermore, we demonstrate how an institution's internal data or program-level data can be combined into the analysis in order to further drill down into the data and make additional comparisons if desired (such as those we conducted by gender). This pilot also illustrates one viable method for determining the types of research (translational level (T-codes)) a program or hub supports and demonstrates how these data can be further combined with bibliometrics and internal data records to evaluate scholars' early outcomes.

Step 1: Using Thomson Reuters to Access Bibliometric Information on All ICTR Publications

An Excel file of publications that cited ICTR (and thus used ICTR resources) from 2008 to 2015 was sent to Thomson Reuters so that bibliometric data on ICTR publications could be compiled and linked to each publication. Of the list of 2153 publications sent to Thomson Reuters, bibliometric data were available for 1917 publications (89%) from Web of Science (519 of which were KL2 publications). Thomson Reuters supplied the results in the form of both an Excel spreadsheet of publication-level data and a Microsoft Access database file that contained tables of authors, publications, and citing and cited publications. The Access database was intended to be used with the supplied Xite software for creating reports and database manipulation, however we discovered that the Xite software did not enable us to drill down into the data and make the sort of comparisons we desired. As a result, tables from the Access database supplied by Thomson Reuters were individually exported into Excel and subsequently imported into SPSS and merged together using a combination of PMID, Author and

ISI LOC (which is a unique article identifier that is assigned to an article when it enters Thomson Reuter's database).

Step 2: Merging in Author-Level Data

In order to account for the numerous authors associated with each publication, the publication-level data for the 1917 ICTR publications were merged together with the author-level bibliometric data tied to the 7051 unique authors listed on the ICTR publications. The resulting data file of 14,030 cases links both publication and author-level bibliometric data together by creating a separate record for each author by each publication. This means that each publication or unique PMID is listed multiple times (once for each individual author listed on the publication). This format was used to account for the multiple authors who contributed to a publication and preserve the data on author citation order and the author-level bibliometric data associated with each individual author.

Merging in author-level data poses its own unique set of complex issues, sometimes referred to as the "author name disambiguation" problem. This refers to the ambiguity that plagues linking together author information because author names are ultimately not unique enough to avoid duplication and because of the inconsistent manner in which author names are reported [8]. Author name ambiguity is a commonly acknowledged issue, one that has been the subject of multiple publications and disambiguation methods of varying complexities. The problem has given rise to alternatives for tracking author publishing activity such as the use of ORCID which utilizes an author ID system [8]. Considering these author name disambiguation issues, we manually inspected bibliometric records using data files containing KL2 scholars' demographic and academic profile traits. We used these data to aid in disambiguation of KL2 scholars' author-level bibliometric data and to merge demographic information on the KL2 scholars into the SPSS database. Merging internal data records containing data on scholars' demographic and academic profile traits with bibliometric data allows us to examine the potential influence of gender on scholars' productivity and the impact of their research. Gender as a variable of interest was included in this study in part because previous studies have reported significant gender disparities in traditional measures of career success such as publication productivity [9].

Two different sources provided information on KL2 scholars' demographic and academic profile traits. Demographic data exported from our local tracking system, WEBCAMP, were used in conjunction with internal ICTR KL2 Excel sheets containing verified data on scholar background and careers. Using these two different sources, we crafted a demographic datafile on KL2 scholars that contained information on KL2 scholars' gender, race/ethnicity, department and division, career outcomes (specifically, career position/titles), continued involvement in research, and retention at UW-Madison. We used the information from the demographic datafile to review the bibliometric SPSS datafile case by case and identify the exact author name iterations that match the KL2 scholars' bibliometric data obtained from Thomson Reuters. We next merged the KL2 scholars' demographic and academic profile data with the Thomson Reuter's bibliometric file using author name to link and merge the 2 data sets. Using the data from the newly merged SPSS datafile, we identified 519 KL2 publications. For the purposes of this pilot, KL2 publications are defined as any ICTR publication on which at least 1 KL2 scholar is listed as an author (either as a result of their KL2 participation or their use of other ICTR services). We used this list of 519 KL2 publications and their associated PMIDs to begin the process of identifying the types of research produced by the KL2 scholars.

Step 3: T-Coding and Classifying Publications Along the Translational Spectrum

To examine the types of research the KL2 program supports, we first needed to code or classify each scholar's publications along the

translational spectrum. To accomplish this, we used a checklist to manually classify publications into translational research categories. The checklist used is described in Surkis *et al.* [5]. It uses categories from the 2013 IOM report as a starting point on which agreed-upon definitions across 5 CTSA institutions were developed. This conceptualization of the translational spectrum is made up of 5 categories that were further simplified into 3 categories of publications. These are T0 (basic science research), T1–T2 (clinical research), and T3–T4 (post-clinical translational research). An additional category called "TX" was used for publications deemed to be outside the translational spectrum.

The coding process involved 3 different coders, each of whom individually reviewed the abstracts and MESH terms of the 519 publications and assigned each publication to a category along the translational spectrum accompanied by the checklist number that led them to choose that specific translational category [5]. Discrepant codes were then discussed by the coding group to arrive at a final consensus code. Publications where coders' individual category assignments differed and no final consensus coding could be reached after discussion were sent to a coder outside the CTSA institution (one of the authors on the original article from which this checklist was crafted, Surkis *et al.* [5]) for final coding input. An Excel file containing the final hand-coded T-codes for the 519 KL2 publications was then merged into the SPSS data set using PMIDs as the key field.

Step 4: Construction of Key Analytic Variables: Calculating Normalized Citation Indicators or Comparative Citation Ratios (CCR)

In addition to the bibliometric data supplied by Thomson Reuters, we also included data from other sources such as iCite and the Essential Science Indicators (ESI) in order to craft normalized citation impact indicators for comparison purposes. These are what Schneider *et al.* [10] have called CCR. CCR indicators compare the observable citation rates for a set of publications with the citation rates of another "comparable" set of publications, so that publications from different fields and years can be compared with one another [10]. CCR indicators are ultimately a simple ratio created by taking the observed citation rate divided by the expected citation rate. Different CCR indicators vary in how the expected citation rate is calculated and therefore what is chosen as a suitable "comparable" set of similar publications [10]. An additional benefit of these CCR indicators is that as a result of taking the form of a ratio the results are easy to interpret, where 1.00 equals a publication (or group of publications) that have received the exact number of citations that would be expected. A score above 1.00 indicates a performance better than expected, and a score below 1.00 represents performance lower than expected.

Two such CCR indicators were created using data exported from the ESI database provided by Thomson Reuters. ESI assigns papers to 1 of 22 different fields, providing field baseline data in the form of annual expected citation rates that can be used to assess the impact of papers against a field-appropriate benchmark. An article's assigned field is determined by the journal in which an article appears [11]. Using the master journal list provided by Thomson Reuters, it is possible to link Web of Science publication data with ESI baseline data by matching articles based on the year and journal in which they were published with the appropriate ESI field [12].

Combining these different data sources allows for normalized CCRs to be calculated in which the number of citations a publication has received is benchmarked against a world average obtained for articles belonging to the same field and published in the same year. Using the matched and merged ESI baseline data, 2 different indicators were created by using 2 different methods of normalization. The first

indicator, Citation Impact, is derived when normalization is done at the group level and is calculated by taking the sum of citations for the group of articles under consideration divided by the sum of the ESI baselines/annual expected citation rates for these articles [11]. The second indicator, Relative Impact, is calculated when normalization is conducted at the individual article level, where an individual citation impact ratio is calculated separately for each article, and normalization is ultimately achieved by calculating the mean of the group's individual citation impact ratios [11].

Step 5: Merging in iCite Data

Additionally, after entering the write-up phase of this pilot, we became aware of the CCR variable called the relative citation ratio (RCR) developed by researchers at NIH and available for free through iCite (<https://icite.od.nih.gov/>) [13]. iCite's RCR variable differs from similar field-normalized CCR variables in that, unlike with Thomson Reuters' approach using their CCR variable ratio, an article's field is not determined by the journal in which an article is published, but rather by creating a field based on an article's co-citation network [13].

The expected citation rate for iCite variables is not taken from a global average but instead is only related to publications produced with NIH funding, making it of particular interest for CTSA hubs.

By uploading into iCite the same publication list of 1917 publications on which we had received bibliometric data from Thomson Reuters, we were able to download the bibliometric analyses done through iCite into Excel. We again used PMIDs to merge the iCite data into our SPSS database. The bibliometric results from iCite services are not directly comparable with those from Thomson Reuters since the bibliometric analysis through iCite was done over a year later and therefore had more time to gain citations. However, we decided to include the metrics from iCite in our analysis (using the same 2008–2013 timeframe we used for the Thomson Reuters bibliometrics) because the RCR has been mentioned as a possible variable in future CTSA common metrics. The full list of bibliometric variables we chose to report on after reviewing the various metrics we crafted and/or pulled using the data from Thomson Reuters, iCite and the ESI database are listed in Table 1 along with metrics descriptions/definitions.

Table 1. Metric descriptions

Metric	Description	Source of metric descriptions
Author-level metrics		
Total number of citations	Total number of cites received by this author	[14]
Number of articles	Number of articles author has published in this data set	[14]
Average citations per paper	Average number of cites per paper	[14]
C-index	The sum of the actual citations divided by the sum of the expected citations equals the crown index (C-index). The C-index can be used to gauge the performance of a body of work, for example, an author's works, in relation to the norms of its constituent journals	[14, 15]
H-index	Is a distribution-based indicator that corresponds to the number of papers at or above a given citation level equal to the value of the citation threshold. This measure attempts to reflect both productivity (number of papers) and impact (number of citations) in 1 number. At the author level, it is considered to be an indicator of a researcher's lifetime scientific achievements	[15]
Average percentile	Is an average of the percentile measurements for the papers in this set. This will tell you how this group of papers has performed overall, with each paper's citation count having been normalized for its field of study and year (indexed year). In the case of an author's body of work, the average percentile would indicate the relative performance of the author's works, having been normalized for whichever fields they fall into	[14, 15]
Self-citations	Number of self-citations	[14]
Publication-level metrics		
Total number of citations	Total number of cites received by this article	[14]
Percentile	The average number of citations received by a paper published in the same journal, in the same year (indexed year), and of the same document type (article, note, review, editorial, etc.)	[14]
Ratio	The ratio of actual citations received by the paper to expected citations for peer papers; calculated by dividing cites by expected cites	[14]
ESI relative impact	A CCR indicator that is calculated when normalization is conducted at the individual article level, where an individual citation impact ratio is calculated separately for each article, and normalization is ultimately achieved by calculating the mean of the group's individual citation impact ratios	[11]
ESI citation impact	A CCR indicator that is derived when normalization is done at the group level and is calculated by taking the sum of citations for the group of articles under consideration divided by the sum of the ESI baselines/annual expected citation rates for these articles	[11]
Author citation order	A sequential number that is given to each author as it appears within the article	[14]
iCite publication-level metrics		
Relative citation ratio	A citation-based measure of scientific influence of one or more articles. It is calculated as the cites/year of each paper, normalized to the citations per year received by NIH-funded papers in the same field and year	[13]
NIH percentile	Percentile rank amongst NIH-funded publications	[13]
Total citations	Total citations found in the citation database	[13]
Citations per year	Citations per full calendar year after publication, through the end of their records	[13]

ESI, Essential Science Indicators; CCR, comparative citation ratios; NIH, National Institutes of Health.

Analysis

The compiled data were analyzed to examine:

- (1) Overall frequencies for publication and author-level metrics for KL2 publications.
- (2) Demographic and academic profile of KL2 scholars including diversity in terms of gender and types of research.
- (3) Degree to which KL2 scholars' demographic and academic profile traits relate to variation in metrics measuring scholarly impact and research productivity.
 - (a) Using within-group KL2 comparisons by gender.
- (4) T0–T4 codes that indicate where on the translational spectrum KL2 scholars' publications fall.

Shapiro-Wilk tests revealed that the data were not normally distributed; as a result, non-parametric tests were used (Mann-Whitney *U* test). All analyses were performed using SPSS version 23. For analyses that utilized bibliometric data, findings are briefly summarized in the text with more detailed information located in the tables (including means, medians, standard deviations, test statistics, and statistical significance/*p* values). For all analyses, statistical significance of α of 0.05 was used; in the tables significant results are indicated by an asterisk*.

Results

Demographics of KL2 Scholars Active Between 2008 and 2013

Of the 58 KL2 scholars and their associated publications for which Thomson Reuters supplied bibliometric data, only publications published between the years 2008 and 2013 were included in the citation analyses (in order to allow publications at least 2 y to gain citations). The 50 KL2 scholars that published between 2008 and 2013 and their associated publications are thus the primary focus of this pilot study. Analyses using T-codes that indicate the types of research being produced by classifying where a publication falls along the translation spectrum, used the full set of KL2 publications from 2008 to 2015.

The gender distribution of the 50 KL2 scholars whose publications were used in citation analyses was nearly even, with 24 scholars (48% of the sample) identifying as female and another 26 scholars identifying as male (52%). The majority of KL2 scholars ($n = 37$) identify as non-Hispanic white (74%). The vast majority of KL2 scholars (94%) were still active in research and 82% still worked at UW-Madison in some capacity as of December 2015. KL2 scholars have a wide variety of areas of expertise and come from many different departments. Most KL2 scholars (for which data were available) come from UW-Madison's School of Medicine and Public Health ($n = 32$) which has the largest pool of potential scholars of all the health sciences schools/colleges.

KL2 Within-Group Analysis: Author-Level Bibliometrics—By Gender

Within-group analysis on KL2 scholars' demographic and academic profile traits was conducted to examine if KL2 author-level bibliometric values differ significantly depending on scholars' traits. We chose to use gender to compare and contrast the performance of our KL2 scholars since many past studies have found that males' and females' publication productivity is significantly different, with females on average publishing less than their male counterparts [9]. Additionally, scholars receiving K awards from NIH have in the past been referred to as "an ideal population within which to study issues of gender and academic success" [16].

Analysis by gender using the Mann-Whitney *U* test indicated that no statistically significant differences in author-level bibliometrics exist by gender (Table 2). Despite the lack of significant differences at the author level, the overall results suggest that the KL2 scholars appear quite productive with the average KL2 scholar publishing in the double digits (median = 10) at ICTR during the 5-year timeframe that was the focus of our bibliometric analysis (2008–2013). Additionally, the Crown index (C-index) which can be used to measure the performance of an author's body of work (by comparing it to the norms of the journals in which they publish) illustrates that KL2 scholars perform above 1.00 and therefore their body of work performs slightly better than expected (median = 1.07). The average percentile that at the author level reflects the relative performance of the author's work normalized to their respective field(s) is also encouraging. Percentile metrics supplied by Thomson Reuters are normalized indicators where lower values indicate better performance; KL2 scholars appear to be performing better than

Table 2. Author-level bibliometrics by gender

KL2 scholars: author-level bibliometrics—by gender	Male (n = 26)	Female (n = 24)	Total (n = 50)
Total number of citations	196.23	149.21	173.66
Mean			
Median	59.00	72.00	71.00
Standard deviation	301.44	266.38	283.25
<i>U</i> statistic		301.00	
<i>p</i> value		0.83	
Number of articles			
Mean	11.23	10.54	10.90
Median	10.00	10.00	10.00
Standard deviation	8.48	9.11	8.71
<i>U</i> Statistic		291.50	
<i>p</i> value		0.69	
Average citations per paper			
Mean	12.48	15.57	13.96
Median	7.78	7.37	7.48
Standard deviation	11.92	17.73	14.92
<i>U</i> Statistic		293.50	
<i>p</i> value		0.72	
C-index			
Mean	1.17	1.36	1.26
Median	1.14	1.03	1.07
Standard deviation	0.54	0.92	0.75
<i>U</i> statistic		299.00	
<i>p</i> value		0.83	
H-index			
Mean	5.12	4.78	4.96
Median	4.00	4.00	4.00
Standard deviation	3.75	4.12	3.89
<i>U</i> statistic		276.00	
<i>p</i> value		0.81	
Average percentile			
Mean	39.51	38.75	39.14
Median	41.46	37.57	40.91
Standard deviation	19.74	21.17	20.23
<i>U</i> statistic		286.00	
<i>p</i> value		0.98	
Self-citations			
Mean	11.60	12.65	12.10
Median	4.00	7.00	5.00
Standard deviation	18.85	29.44	24.24
<i>U</i> statistic		278.50	
<i>p</i> value		0.85	

many of their peers who are publishing within the same field and timeframe.

KL2 Within-Group Analysis: Publication-Level Bibliometrics—By Gender

When analysis by gender at the publication level was conducted, author citation order emerged as the sole significant gender difference. Author citation order refers to where an author falls in a publication's list of authors, with results indicating that female scholars (median = 2) rank significantly higher in their publications citation order than their male counterparts (median = 3), $U = 12,117$, $p < 0.001$. Full results comparing male and female KL2 scholars' publication-level bibliometrics appear in Table 3. Overall results regarding KL2 scholars' publications regardless of gender are encouraging. Thomson Reuter's CCR variable ratio shows that KL2 scholars' papers are performing as expected. The percentile values supplied by Thomson Reuters are also favorable, keeping in mind that percentiles from Thomson Reuter's range from 0 to 100, where lower values indicate better performance. Percentile essentially indicates the percentage of papers (from the same field, year, and document type) that are cited more often than a specific paper or set of paper(s) [14]. Therefore, the low percentiles received by KL2 scholars, regardless of gender, shows that their papers are performing better—that is, cited more often—than the

majority of papers published within the same field, year and of the same document type.

iCite Data: KL2 Within-Group Analysis—By Gender

Bibliometrics downloaded through the free NIH iCite service were also analyzed to see if significant differences existed by gender. Shapiro-Wilk tests revealed that the data were not normally distributed, as a result, analysis by gender was done using Mann-Whitney U tests. No significant differences by gender emerged but results are summarized briefly below as well as in Table 4. While no significant differences in the RCR exist by gender, it is important to note that both male (median = 1.25) and female KL2 scholars' (median = 1.27) RCR is above 1.00 indicating that both genders' publications are cited more often than the norm for NIH publications. NIH percentile data also suggests that KL2 scholars' publications are cited slightly more often than the norm for NIH publications.

KL2 Hand-Coded T-Codes

Hand-coded T-codes indicating where a publication falls along the translational spectrum (T0, T1–T2, or T3–T4) provide another avenue of analysis. Since this area of analysis did not involve citation analysis, it was not necessary to limit the analysis to publications published between 2008 and 2013. All 519 publications authored by the 58 KL2 scholars who published between 2008 and 2015 were coded. The inclusion of the works of the 8 additional scholars and publications from the years 2014 and 2015 did not significantly alter the demographics of the groups (which experienced fluctuations $\leq 2\%$).

After removing the 46 KL2 publications that were coded as TX and deemed to be outside of the translational spectrum, we analyzed the remaining 473 KL2 publications. Overall, KL2 publications are primarily concentrated in either the T0 or T3–T4 categories on

Table 3. Publication-level bibliometrics by gender

KL2 scholars: publication-level bibliometrics—by gender	Male (n = 196)	Female (n = 172)	Total (n = 368)
Total number of citations			
Mean	23.66	19.91	21.91
Median	10.00	10.00	10.00
Standard deviation	44.46	25.30	36.77
U statistic		16,730.50	
p value		0.90	
Percentile (citation percentile)			
Mean	35.39	35.44	35.41
Median	30.93	27.97	28.93
Standard deviation	29.15	29.28	29.17
U statistic		15,672.50	
p value		0.84	
Ratio (actual to expected citations)			
Mean	1.36	1.25	1.31
Median	0.96	0.99	0.99
Standard deviation	1.26	1.06	1.17
U statistic		16,422.50	
p value		0.67	
ESI relative impact			
Mean	1.59	1.46	1.53
Median	0.79	0.80	0.80
Standard deviation	2.41	1.72	2.12
U statistic		16,726.50	
p value		0.98	
ESI citation impact			
Mean	1.74	1.48	1.62
Author citation order*			
Mean	4.37	2.88	3.67
Median	3.00	2.00	2.00
Standard deviation	3.88	3.25	3.67
U statistic		12,117.00	
p value		0.00	

Significant results at $p < 0.05$ are marked with an *.

Table 4. iCite data: KL2 publication-level bibliometrics by gender

iCite—bibliometrics by gender	Male (n = 196)	Female (n = 172)	Total (n = 368)
iCite—relative citation ratio			
Mean	2.32	1.91	2.13
Median	1.25	1.27	1.26
Standard deviation	2.91	1.89	2.49
U statistic		16,384.00	
p value		0.64	
iCite—NIH percentile			
Mean	56.83	55.70	56.30
Median	58.50	59.05	58.60
Standard deviation	29.40	29.16	29.25
U statistic		16,384.50	
p value		0.64	
iCite—total citations			
Mean	32.81	26.31	29.77
Median	15.00	14.00	14.00
Standard deviation	55.16	33.99	46.54
U statistics		16,023.00	
p value		0.41	
iCite—citations per year			
Mean	5.53	3.97	4.80
Median	2.75	2.50	2.63
Standard deviation	7.64	4.30	6.34
U statistics		15,217.00	
p value		0.11	

Table 5. KL2 publication-level—translational level (T-codes) by gender

	KL2 T-codes			Total
	T0 (Basic Science Research)	T1/T2 (Clinical Research)	T3/T4 (Post-Clinical Translational Research)	
Male				
Count	161	32	45	238
Percent	67.6	13.4	18.9	100.0
Female				
Count	64	20	151	235
Percent	27.2	8.5	64.3	100.0
Total				
Count	225	52	196	473
Percent	47.6	11.0	41.4	100.0

opposite sides of the translational spectrum, with the largest category being T0 (basic research) and with very few publications falling into the T1–T2 (clinical research) category.

KL2 Within-Group Analysis: Publication-Level T-codes—By Gender

When looking at T-codes by KL2 scholars' demographic traits, a significant association emerged between T-codes (where research falls along the translational spectrum) and gender ($\chi^2 = 101.9$, $p < 0.001$). Intriguingly, the results indicate that male and female scholars tend to publish articles on different ends of the translational spectrum (see Table 5).

Table 6. Translational level (T-codes) by year

	KL2 T-codes			Total
	T0 (basic science research)	T1/T2 (clinical research)	T3/T4 (post-clinical translational research)	
2008				
Count	16	1	7	24
Percent	66.7	4.2	29.2	100.0
2009				
Count	17	8	15	40
Percent	42.5	20.0	37.5	100.0
2010				
Count	26	7	18	51
Percent	51.0	13.7	35.3	100.0
2011				
Count	33	8	22	63
Percent	52.4	12.7	34.9	100.0
2012				
Count	36	4	40	80
Percent	45.0	5.0	50.0	100.0
2013				
Count	43	10	28	81
Percent	53.1	12.3	34.6	100.0
2014				
Count	30	10	43	83
Percent	36.1	12.0	51.8	100.0
2015				
Count	24	4	23	51
Percent	47.1	7.8	45.1	100.0
Total				
Count	225	52	196	473
Percent	47.6	11.0	41.4	100.0

The majority (67.6%) of male-authored KL2 publications were classified as T0 (basic science publications) compared with their female counterparts for whom T0 publications made up a minority (27.2%) of their publications. KL2 scholars authored few T1–T2 publications; for both genders the smallest percentage of their publications fell within the T1–T2 category. Overall however, a larger percentage of male KL2 scholars' publications fell within the T1–T2 category (13.4%) compared with their female counterparts (8.5%). However, female scholars published far more T3–T4 publications than their male counterparts. The majority (64.3%) of female KL2 scholars' publications were T3–T4 publications compared with male scholars for whom only 18.9% of their publications fell within the T3–T4 category. The distribution of each gender's publications along the translational spectrum illustrates that male scholars' research is heavily concentrated towards the early/basic science research end of the spectrum, while female scholars research tends towards the T3/T4 translational end of the spectrum.

The observable differences in publication T-codes between the KL2 scholars are likely at least in part due to other factors. For instance, changes over time may be a factor. An analysis of T-codes by year seen in Table 6 illustrates an overall trend towards a greater proportion of KL2 scholars' research falling within the realm of T3/T4 research over time.

Discussion

Early outcomes suggest that KL2 scholars are achieving extrinsic career success in terms of research and publications. Our results show that KL2 scholars' publications are cited more often than the norm for NIH publications. In this study, we used several sources of bibliometric data providing several normalized citation impact indicators or CCR to compare across fields and time. Despite the different databases used to produce these indicators, differences in how these CCR indicators are calculated, and the different points in time when these data were pulled, the results were similar and illustrated that the KL2 scholars' publications are generally performing either as expected or better than expected. We also observed favorable results and returns from investing in the next generation of researchers in the number of scholars that have continued to work for the institution (82%) and stayed engaged in research (94%). This level of research engagement matches the productivity reported by Schneider *et al.* [2] and is similar to that reported by Amory *et al.* [17] on other K scholars at another large public institution (in the top 10). Additionally, we found outcomes that suggest some success in developing independent researchers, for instance, 84% of our KL2 scholars published at least one publication as a first author which is higher than the 72% reported in a previous study by Schneider *et al.* [2]. Overall, several of our findings support those of previous studies on K scholars, further adding to the evidence of positive early outcomes from scholars supported by the KL2 program.

The distribution of publications across the translational spectrum is of considerable interest and importance. The CTSA program—"clinical and translational science awards"—was originally intended to speed the translation of basic research (already being conducted in abundance with existing NIH funding) to clinical and practice applications. However, our data (Table 5) indicate that 47.6% of the research conducted by scholars in the KL2 program is at the basic science research level, and only 11% is at the clinical level (with 41.4% at the T3/T4 level). This raises the intriguing question of whether the *clinical* research infrastructure is adequately supported by the CTSA program. After all, according to the most recent funding announcements, the CTSA program is no longer supporting clinical research infrastructure as it did in its earliest years. While we do not know with certainty what the distribution of T0–T4 research is at other CTSA hubs, anecdotal evidence suggests our institute's experience may not be unique. To the extent that is the case, the large number of publications we observed at the basic science research level (T0), and the relatively small number at the T1–T2 level, raises questions as to whether CTSA funding is in fact

fulfilling the mission of enhancing *clinical* and translational research by supporting research across the entire translational spectrum.

In addition to evidence of positive outcomes from scholars supported by the KL2 program, our analyses also produced interesting findings regarding gender and T-codes. For instance, the KL2 program has experienced a shift over time in the distribution of publications across the translational spectrum. This shift towards more T3/T4 work in recent years suggests that concentrated efforts (by the institution and program staff) to support more T3/T4 work have been successful, and also underscores the value of tracking the type of research produced by a program, and by extension a CTSA hub. Understanding the type of research that is being supported and produced is an important first step that allows for the opportunity to make impactful changes with concentrated efforts. For instance, by admitting more scholars whose background and research interests are concentrated in the portion of the translational spectrum that a program or institution may want to prioritize, this small change and shift in priorities can potentially have measurable and significant effects on the stage and type of research that not only the KL2 program produces but by extension the CTSA hub as well. Due to the sheer number of publications KL2 scholars produce that represent roughly 25% of ICTR's publications, shifting the type/stage of research generated by some of ICTR's most productive researchers allows for major shifts to be made in the type of research supported by a CTSA hub as a whole.

Another interesting finding was the limited number of significant differences by gender observed in this study. The lack of significant gender differences in research productivity and impact in this article is somewhat surprising considering the amount of past literature reporting significant gender disparities between men and women in citation behavior and research productivity, and the significant gender differences also reported in the past even among K award recipients [16]. Our analyses found only small non-significant differences between male and female scholars' publication impact and productivity, with the only significant gender differences being observed in author citation order and T-codes.

The significance (if any) behind the gender differences observed between male and female scholars in author citation order and their tendency to publish articles on different ends of the translational spectrum could be the result of a host of factors. Potential explanations include changes over time (such as the institution's increased emphasis on more T3/T4 work in recent years), differences in citation behavior in and across fields, or differences in collaborative practices. Another possible explanation for these gender differences is the different areas of study/expertise that KL2 women and KL2 men chose to study and research. For instance, past research has found that men and women specialize in different areas of research [18], and while most KL2 scholars came from UW-Madison's School of Medicine and Public Health, the KL2 women represent a wider variety of disciplines and areas of expertise than the men.

Overall, further research is needed to better interpret the results observed in our study to see if these findings are generalizable to the KL2 program as a whole or are simply characteristic of the one unique CTSA hub that was the focus of our analysis. Additional work is also needed to discern if the lack of observed gender differences between male and female KL2 scholars in regard to research productivity and the impact of their respective publications reflects identifiable individual, program or institutional level factors that narrow productivity and career trajectory disparities.

Overall, NIH-supported early-career development programs could benefit immensely from more systematic evaluation aimed at understanding the impacts of these programs on the early-career success of scholars. A greater understanding of the multiple factors affecting career success could help career development programs to better

allocate resources and make effective data-driven decisions regarding program changes and improvements.

Using bibliometric data to evaluate CTSA research productivity offers a promising contribution to a broader evaluation of the KL2 program. CTSA evaluators have recommended informally to NCATS that a consortium-level analysis of the KL2 scholars, their programs, and their research productivity would be valuable to the roughly 60 institutions participating in the CTSA consortium. A focus on publications, bibliometrics and translational level for scholars' productivity, across the entire CTSA consortium, could provide consistent in-depth insight into trends related to workforce development objectives in biomedical research in this country.

Two limitations of this study, including limited data and lack of an appropriate comparison group, could be addressed (at least in part) by a consortium-level analysis of the KL2 scholars or by facilitating evaluator access to existing NIH data. For instance, within the scope of this study, we did not include data on current and former KL2 scholars' grant applications and funding history which could have been a useful additional outcome to consider in examining scholars' early career outcomes. Additionally, due to limited data, a suitable comparison group for the KL2 scholars comprised of other early-career researchers from similar fields was not available. Comparisons were instead limited to within-group comparisons and comparisons using various normalized citation impact indicators or CCR. These were useful to compare KL2 scholars' publications to NIH publications via iCite data or to norms calculated using ESI data (wherein publications are assigned to a particular field by journal). If more data were available, a comparison group could potentially have been created and additional analyses undertaken. A consortium-level analysis of the KL2 scholars, in addition to providing interesting insights, would provide a benchmark against which we and other CTSA hubs could gauge the success of our KL2 scholars and evaluate our individual KL2 programs. In addition, we agree with comments made by Schneider *et al.* [2] about the evaluative benefits of having greater access to existing NIH data such as the NIH Information for Management, Planning, Analysis and Coordination database which houses data about career development award applicants, including NIH grant applications and demographic information.

Ultimately our study points to the feasibility of using program-level data and publication data, including bibliometrics and data categorizing publications along the translational spectrum, to evaluate early outcomes including the productivity of scholars and the influence of their published research. Our pilot has illustrated the feasibility of 3 different approaches to bibliometrics: utilizing paid services or the free downloadable reports/datafile options through providers such as Thomson Reuters or Elsevier, or using the free services offered by iCite. Each has advantages and disadvantages (depending on one's needs, goals, and resources) but all are viable options for evaluative purposes. Our study also illustrates the evaluative benefits of categorizing publications along the translational spectrum to identify the types of research produced by a program and its scholars and how these data can be further combined with internal data records and bibliometrics to gain richer insights.

Future Directions

This pilot study illustrates that linking bibliometric data and data categorizing publications along the translational spectrum, with a CTSA hubs' internal program data records is feasible, and offers several innovative possibilities for the evaluation of a CTSA as a whole as well as a targeted evaluation of its programs and investigators. Future directions include building on the existing bibliometric data sets and including the publications of KL2 scholars from other CTSA hubs to provide more context for how our CTSA and KL2 scholars are performing. Another option includes expanding our focus and gathering our KL2 scholars' full publication records to compare how

scholars' publications before, during and after the KL2 program fare. Other avenues of interest include going beyond examining one single aspect of extrinsic career success (publications) and instead combining the quantitative data analyzed within the scope of this study, with KL2 grant/funding data, survey data and qualitative interview data. Efforts are already underway to pursue a study focused on outcomes from KL2 scholars using a mixed-methods approach that builds on the work done by Rubio *et al.* [3] and Lee *et al.* [4] regarding evaluating comprehensive career success by combining our existing data with survey data and data from interviews with KL2 scholars. Utilizing a mixed-methods approach integrating qualitative and quantitative data (e.g., Hogle and Moberg [19]) will allow us to gain a more in depth understanding of career success, and explore the influence that various factors and barriers can have on career direction, research productivity, and even the types of research scholars undertake.

Acknowledgments

The authors thank Linda Scholl for her help in translational spectrum coding and her feedback in reviewing this manuscript. The authors also thank Alisa Surkis for her help in providing translational spectrum coding input on KL2 publications. This work was funded in part, by NIH NCCATS grant UL1TR000427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosures

The authors have no conflicts of interest to declare.

References

1. **Reis SE, et al.** Reengineering the National Clinical and Translational Research Enterprise: the strategic plan of the National Clinical and Translational Science Awards Consortium. *Academic Medicine* 2010; **85**: 463–469.
2. **Schneider M, et al.** Developing the translational research workforce: a pilot study of common metrics for evaluating the clinical and translational award KL2 program. *Clinical and Translational Science* 2015; **8**: 662–667.
3. **Rubio DM, et al.** A comprehensive career-success model for physician-scientists. *Academic Medicine* 2011; **86**: 1571–1576.
4. **Lee LS, et al.** Clinical and translational scientist career success: metrics for evaluation. *Clinical and Translational Science* 2012; **5**: 400–407.
5. **Surkis A, et al.** Classifying publications from the clinical and translational science award program along the translational research spectrum: a machine learning approach. *Journal of Translational Medicine* 2016; **14**: 235.
6. **Leshner AI, et al. (eds)** *The CTSA Program at NIH: Opportunities for Advancing Clinical and Translational Research*. Washington, DC: The National Academies Press, 2013.
7. **Coryn C.** The use and abuse of citations as indicators of research quality. *Journal of MultiDisciplinary Evaluation* 2007; **3**: 115–121.
8. **Milojevic S.** Accuracy of simple, initials-based methods for author name disambiguation. *Journal of Informatics* 2013; **7**: 767–773.
9. **Dehdarirad T, Villarroya A, Barrios M.** Research on women in science and higher education: a bibliometric analysis. *Scientometrics* 2015; **103**: 795–812.
10. **Schneider M, et al.** Feasibility of common bibliometrics in evaluating translational science. *Journal of Clinical and Translational Science* 2017; **1**: 45–52.
11. **Van Veller M, et al.** *Bibliometric Analyses on Repository Contents for the Evaluation of Research at Wageningen UR*. Singapore: World Scientific, 2009.
12. **Thomson Reuters.** Master Journal List [Internet], 2016 [cited Dec 7, 2016]. (<http://ipscience-help.thomsonreuters.com/incitesLive/ESI/Group/overviewESI/scopeCoverageESI.html>).
13. **Hutchins BI, et al.** Relative Citation Ratio (RCR): a new metric that uses citation rates to measure influence at the article level. *PLOS Biology* 2015; **14**: e1002541.
14. **Thomson Reuters.** *Xite 7.2 User Guide*. Philadelphia, PA: Thomson Reuters, 2010.
15. **Thomson Reuters.** *Using Bibliometrics: A Guide to Evaluating Research Performance with Citation Data*. Philadelphia, PA: Thomson Reuters, 2008.
16. **Jagsi R, et al.** Similarities and differences in the career trajectories of male and female career development award recipients. *Academic Medicine* 2011; **86**: 1415–1421.
17. **Amory J, et al.** Scholarly productivity and professional advancement of junior researchers receiving KL2, K23, or K08 awards at a large public research institution. *Journal of Clinical and Translational Science* 2017; **1**: <https://doi.org/10.1017/cts.2016.22>.
18. **Grant L, Ward KB.** Gender and publishing in sociology. *Gender and Society* 1991; **5**: 207–223.
19. **Hogle J, Moberg DP.** Success case studies contribute to evaluation of complex research infrastructure. *Evaluation & the Health Professions* 2014; **37**: 98–113.