Evaluating the research domain and achievement for a productive researcher who published 114 sole-author articles

A bibliometric analysis

Mei-Yuan Liu, MSa,b,c, Willy Chou, MDc,d, Tsair-Wei Chien, MBAe,∗, Shu-Chun Kuo, MDf,g, Yu-Tsen Yeh, MS, Po-Hsin Chou, MD, PhDj,k

Abstract

Background: Team science research includes authors from various fields collaborating to publish their work on certain topics. Despite the numerous papers that discussed the ordering of author names and the contributions of authors to an article, no paper evaluated (1) the research achievement (RA) and (2) the research domain (RD) for productive sole-author researchers.

In addition, few researchers publish academic articles without co-author collaboration. Whether the bibliometric indexes (eg, h-/x-index) of sole-author researchers are higher than those of other types of multiple authors is required for comparison. We aimed to evaluate a productive author who published 114 sole-author articles with exceptional RA and RD in academics.

Methods: By searching the PubMed database (Pubmed.com), we used the keyword of [Taiwan[affiliation]] from 2016 to 2017 and downloaded 29,356 articles. One physician (Dr. Tseng from the field of Internal Medicine) who published 12 articles as a single author was selected. His articles and citations were searched in PubMed. A comparison of various types of author ordering placements was conducted using sensitivity analysis to inspect whether this sole author earns the highest metrics in RA. Social network analysis (SNA), Gini coefficient (GC), pyramid plot, and the Kano diagram were applied to gather the following data for visualization:

(1) the author collaborations and RA using x-index;
(2) the author’s article-related journals frequently published in the past;
(3) the most influential medical subject heading (MeSH) using citation analysis to denote the author’s RD.

Results: We observed that

(1) DR Tseng contributed 114 sole-author articles in 140 publications (≈81.4%) since 2002;
(2) the 100% sole-author scenario earned the highest h-/x-index;
(3) the author’s RD includes epidemiology, complications, and metabolism with an exceptional GC (≈0.55).

Editor: Daryle Wane.
Ethics approval and consent to participate was not applicable.
All data were downloaded from MEDLINE database at pubmed.com.
Consent to publish: Not applicable.
All data used in this study is available in Supplemental Digital Content files.
The authors have no funding and conflicts of interest to disclose.
Supplemental Digital Content is available for this article.
The datasets generated during and/or analyzed during the current study are publicly available.

∗Correspondence: Po-Hsin Chou, 18F, 201, Section 2, Shipai Road, Beitou District, Taipei, 112, Taiwan (e-mail: choupohsin@gmail.com); Tsair-Wei Chien, Department of Medical Research, Chi Mei Medical Center, Tainan, Taiwan (e-mail: smile@mail.chimei.org.tw).

Copyright © 2020 the Author(s). Published by Wolters Kluwer Health, Inc.
This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.
How to cite this article: Liu MY, Chou W, Chien TW, Kuo SC, Yeh YT, Chou PH. Evaluating the research domain and achievement for a productive researcher who published 114 sole-author articles: a bibliometric analysis. Medicine 2020;99:21(e20334).
Received: 11 December 2019 / Received in final form: 27 March 2020 / Accepted: 21 April 2020
http://dx.doi.org/10.1097/MD.0000000000020334
Conclusions: The metrics on RA are high for the sole author studied. The author’s RD can be denoted by the MeSH terms and measured by the GC. The author-weighted scheme is required for quantifying author credits in an article to evaluate the author’s RA. Social network analysis incorporating the Kano diagram provided insights into the relationships between actors (e.g., coauthors, MeSH terms, or journals). The methods used in this study can be replicated to evaluate other productive studies on RA and RD in the future.

Abbreviations: AIF = author impact factor, AWS = an authorship-weighted scheme, GC = Gini coefficient, RA = research achievement, RD = research domain, SCI = scientific citation index, SNA = social network analysis.

Keywords: Gini coefficient, google maps, kano diagram, pyramid plot, research achievement, research domain, sensitivity analysis, social network analysis.

1. Introduction

Team science research is defined as the collaboration of numerous coauthors from various fields join in working toward the resolution of research issues.[1] The number of authors collaborating on an article showed an increasing trend in the past.[2] Collaborative research networks can help other researchers understand the relationship among the members of a science research team.[3] However, several isolated (or sole) authors with numerous publications should be a concern in the authorship network.

Numerous metrics (eg, h-/g-/x-index)[4–6] proposed for assessing individual RA (IRA) are based on the assumption that all coauthors equally contribute to an article, a condition that is notably unfair and unreasonable. Accordingly, numerous counting schemes,[7] such as fractional counting[8–12] and authorship-weighted counting,[10–12] have been addressed for quantifying co-author credits besides the traditional full counting (ie, all authors contributed equally to an article).[13] We were motivated to apply 1 of an authorship-weighted scheme (AWS)[2] to compare the differences in metrics among scenarios with various author positions in an article byline.

The bibliometric analysis is defined as a statistical method used for assessing the academic quality of journals or authors and citation rates. The knowledge structure and development of research fields can be understood based on related publications.[5] As of March 2, 30 articles were searched by term of (“Medicine (Baltimore)”[Journal] AND bibliometric[MeSH Major Topic] in Medicine (Baltimore).[14–18] Bibliometric studies of scientific collaboration have been conducted in various fields,[19,20] providing different levels of cooperation frequency in research practice.[21] One of the methods used to study such collaboration is the co-occurrence network analysis, which focuses on finding patterns of contacts or interactions between social actors (e.g., coauthors, MeSH terms, or journals) as the subjects of co-occurrence relationship. Thus, analyzing co-occurrence relationship can better reflect the truth of scientific research and academic communication,[20] including the research domain (RD) using MeSH terms.[22]

Although numerous papers have discussed the ordering of author names and the contributions of authors to an article,[2,8–13] no paper evaluated

(1) the RA (based on metrics) and
(2) the RD (based on Gini coefficient (GC)[23]) for the productive researchers who published numerous sole-author articles.

The RD median of the GC (0.32) for the middle authors are notably less than those for the first (0.53) and last (or deemed as corresponding) authors (0.42).[22] Whether the RD based on the GC for sole authors is the greatest among all types of author orderings in article positions (ie, >0.33, 0.43, and 0.53 as mentioned above) is worthy of study.

This study aimed to

(1) compare the metrics on RA among different types of author positions in articles and
(2) evaluate 1 author who published a number of sole-author articles with exceptional RA and RD in academics.

2. Methods

2.1. Data source

Two steps were conducted for data organizations. First, we searched the PubMed database (Pubmed.com) using the keyword (Taiwan [affiliation]) from 2016 to 2017 and downloaded 29,356 articles. A total of 371 (1.26%) sole-author and 12 articles (3.23%=12/371) authored by DR Tseng, who worked for the Department of Internal Medicine National Taiwan University College of Medicine, Taiwan, were obtained, see dataset in Supplemental Digital Content file 1, Available at: http://links.lww.com/MD/E269.

Second, we downloaded 140 abstracts from PubMed by searching Tseng, Chin-Hsiao [Author - Full] and 114 (1414/140) sole-author articles.

2.2. Representations of the research results

2.2.1. Pyramid plot for displaying author publications and citations.

The publications and citations for the studied author (DR Tseng) since 2002 were drawn using the pyramid plot.
2.2.2. Contingency table for comparing the RA metrics. A comparison of various types of author ordering placements (from 1 to 4, with 4 as a corresponding author) was performed using sensitivity analysis to inspect whether the sole author obtains the highest RA metrics (ie, h/x-index and author impact factor [AIF]). Two panels based on article citations and the article-based journal impact factors (JIF) were included in this contingency table.

The citation-based x-index was computed by the core publications at i (cai) and the corresponding article citations at cai in descending order (see the definition of x-index[8] in Eq. (1)). Similarly, the Sciences Citation Index (SCI) using x-index was referred to as the article-based JIF provided by Clarivate Analytics in 2019 in Eq. (2).

\[
x_{\text{Citation}_i} = \sqrt{\max(i \times c_{ai})},
\]

\[
x_{\text{SCI}_{i}} = \sqrt{\max(i \times \text{SCI}_{ai})},
\]

The AWS[23] in Eq. (3) was applied to quantify an author’s credits on an article denoted by the author Cai and SCIai, in Eqs. (4) and (5), where m-1 denotes the number of coauthors. The first author denoted by m=0 exhibited the highest contribution, followed by the last (ie, corresponding, m=1) author; other middle authors were denoted by the symbol m from 2 to m-1:

\[
W_m = \frac{\exp(y_m)}{\sum_{m=0}^{m-1} \exp(y_m)} = \frac{(2.27)^m}{\sum_{m=0}^{m-1} (2.27)^m}.
\]

\[
C_{ai} = W_i \times c_i,
\]

\[
\text{SCI}_{ai} = W_i \times \text{SCI}_i.
\]

The citable numbers of articles (=publication output = Np) were computed by Eqs. (4) and (5), where the cai and SCIi for each article were set at 1.0. The h-index is defined as “A scientist has index h if h of his or her Np papers have at least h citations each, and the other (Np – h) papers have fewer than \(\leq h\) citations each”.[4] The AIF is equal to \(\frac{\text{C}_{\text{total}}}{N_p}\).

2.2.3. Author RA and RD

2.2.3.1. Author collaborations and RA. Social network analysis (SNA)[3] was applied to gather the coauthors’ contributions on each article to visualize the relationship among authors. The closely connected actors with high numbers of co-occurrences are linked by a wide line. The bubble size was determined by the number of connections with other actors in the network.

Author x-indexes were dispersed on the Kano diagram,[24] where the core publications and citations were on x and y axes. The bubbles were sized by the x-index.

2.2.3.2. Author publications in journals. The networks constructed by the author and article-based journals were plotted using SNA. Bubble size was determined by the frequency author publications in journals.

2.2.3.3. Author RD. Medical subject headings (MeSH) terms is the NLM controlled vocabulary used to manually index articles for MEDLINE/PubMed.[25] Similar to the author network, MeSH terms were assigned as authors listed in each article. The largest bubble representing an influential role in the subnetwork features the highest number of connections related to other terms. The top 5 representatives were used for computing the GC as performed in a previous study.[22] The higher is the GC, the more distinct is the RD appearance. We will inspect whether the RD GC for the studied sole author is the greatest among all types of author orderings in article positions (ie, \(>0.33, 0.43,\) and \(0.53\)).

The AWS was applied to compute the weights for MeSH terms in each article using the traditional full counting (ie, all authors contributed equally to an article).[11] The parameter \(y_m = 0\) is set in Eq. (3). Using Eqs. (1) and (4), the x-index for each MeSH term can be obtained, through which we can examine the most influential term by using the Kano diagram.

2.2.4. Kano diagram used in this study. The Kano diagram is based on a theory of product development and customer satisfaction developed in 1984 by Professor Noriaki Kano,[24] who classified products or items into 3 main categories of quality: basic requirement, one-dimensional quality, and exciting feature. These items are diagrammed according to the satisfaction perceived by customers on axis Y and the effort achieved by providers on axis X.

2.2.5. Creating dashboards on google maps. The figures were plotted by author-made modules in Excel (Microsoft Corp). We created hypertext mark-up language pages used for Google Maps. All relevant actor information can be linked to dashboards on Google Maps, see Supplemental Digital Content 2, http://links.lww.com/MD/E270.

3. Results

3.1. Pyramid plot for displaying author publications and citations

DR Tseng contributed 114 sole-author articles in 140 publications (=81.4%) since 2002. The highest numbers of publications and citations in these 140 articles were 19 and 370 in 2013 and 2012, respectively (Fig. 1). The grey color bars denoted the weights computed by the AWS in Eq. (3) and (4). The AIF for Dr. Tseng were decreased from 12.5 to 10.9 in comparison to those 2 non-weight and weighted AWS with bars in color.

The most cited article (PMIC = 22889723 cited by 110 articles) published in 2012 is related to the evaluation of the association between arsenic and diabetes.[26]

3.2. Contingency table for comparing the RA metrics

Six scenarios are presented in Table 1. The original data comprised the 140 studied articles. The RA on the 100% sole-author scenario is higher than the other scenarios (Table 1). The AIF is not affected by the AWS in scenarios 2 to 6 due to the equally proportional shares in citable and the cited parts using Eqs. (4) and (5). From the data Table 1, the AWS can solve the problematic assumption that all coauthors equally contribute to an article. The sole author owns the full part of contributions in an article.

3.3. Author RA and RD

3.3.1. Author collaborations and RA. DR Tseng accounted for the largest bubble connected to his colleagues as determined by SNA (Fig. 2; top). Seven clusters are separated by colors. The members closer related to their coauthors are within a similar cluster.
DR Tseng presents the largest bubble with $x$-index = 20.78 ($=\sqrt{48 \times 9}$ in Eq. (1)) at the right-bottom corner of the Kano diagram in Figure 2 (bottom). The RA is highlighted by the core publications ($=48$) and citations ($=9$). By contrast, another author (Elizabeth A. Maull), who is the first author of an article (PMIC = 22889723 cited by 110 articles) published in 2012, is located at the left-top corner in Figure 2 (bottom) and earns the $x$-index = 8.34, the cited = 69.53, the citable = 1, and h-index = 1.

Interested readers are invited to scan the QR code in Figure 2 and click the bubble of interest to view the details about the information of metrics and publications in PubMed.

### 3.3.3. Author RD

As for the author RD using the MeSH terms in Figure 4, the top 3 are epidemiology, complications, and metabolism with an exceptional RD GC ($=0.55$) in Figure 4 (top), greater than the median of all types of author orderings in article positions (ie, $>0.33, 0.43, \text{and} 0.53$).

The most productive MeSH term is epidemiology at the right-bottom corner in Figure 4 (bottom) with $x$-index = 9.57, the cited = 7.64, and the citable = 12 on the x-core publication (ie, the publications at $i$ based on Eq. (1)).

Another influential MeSH term is mortality at the left-top corner in Figure 4 (bottom) with $x$-index = 7.68, the cited = 59, and the citable = 1.

### Table 1

Comparisons in metrics for various scenarios of author ordering placements.

| No | Scenario          | Author | $h$-index | Ci | $k$ | $x$-index | Citable | Cited | AIF |
|----|-------------------|--------|-----------|----|-----|-----------|---------|-------|-----|
| SCI based                          |
| 1  | Original data     | $>=1$  | 7         | 2.6| 84  | 14.85     | 110.27  | 426.6 | 3.87|
| 2  | 100% sole author  | 1      | 7         | 2.8| 98  | 16.45     | 140.00  | 487.2 | 3.48|
| 3  | First author      | 4      | 5         | 1.8| 98  | 13.20     | 90.15   | 313.7 | 3.48|
| 4  | 2nd author        | 4      | 1         | 1.3| 5   | 2.58      | 12.20   | 42.5  | 3.48|
| 5  | 3rd author        | 4      | 0         | 0   | 0   | 0         | 4.48    | 15.6  | 3.48|
| 6  | Last author       | 4      | 3         | 1   | 35  | 5.94      | 33.16   | 115.4 | 3.48|
| Citation based                      |
| 1  | Original data     | $>=1$  | 17        | 9  | 48  | 20.78     | 110.27  | 1198  | 10.9|
| 2  | 100% sole author  | 1      | 22        | 11 | 53  | 24.15     | 140.00  | 1754  | 12.5|
| 3  | First author      | 4      | 17        | 7.1| 53  | 19.38     | 90.15   | 1129  | 12.5|
| 4  | 2nd author        | 4      | 4         | 1  | 48  | 7.08      | 12.20   | 152.9 | 12.5|
| 5  | 3rd author        | 4      | 2         | 1.1| 14  | 3.96      | 4.48    | 56.2  | 12.5|
| 6  | Last author       | 4      | 9         | 2.6| 53  | 11.75     | 33.16   | 415.5 | 12.5|

Publications = 140.

AIF = author impact factor, SCI = scientific citation index.
4. Discussions

4.1. Principal implications from the findings

Team science research with a number of authors collaborating on an article has shown an increasing trend. Few researchers publish academic articles without co-author collaboration. We conducted a small study searching journal articles in Medicine (Baltimore) till 2016. Sole-author articles only account for 0.87% (= 63/7203). No any author was found that she (or he) has two sole-author articles in Medicine (Baltimore). The sole-author representative, DR Tsend with paramount RA and exceptional RD, was particularly illustrated in this study. Interested readers are invited to read the article showing the types of DR Tsend’s publications closely associated with the usage of National Health Insurance Database.

The subsequent questions were thus come up to the sole-author’s RA/RD that can be compared with other individual RAs and RDs. The next research question about the methodology used to report the author’s RA/RD was emerged.

In this study, we observed that

1. DR Tsend wrote 81.4% (=114/140) of sole-author articles since 2002;
2. the AWS effect appears the more number of sole-author articles an author has, the higher RA is possibly earned by the author;
3. the author’s RD can be extracted from the MeSH terms and by the SNA approaches along with the GC denoting the RD depth if GC ≥0.40 or width if GC < 0.40.

4.2. Strengths and features of this study

In tradition, previous studies addressing

1. a high JIF associated with the publication of reviews and original articles instead of those case reports;
2. rigorous systematic reviews receiving more number of citations than other narrative reviews;
3. case reports with low JIF.
Using MeSH clusters as an approach to differentiate article types and citation likelihoods have been proposed in the previous article. Thus far, no such MeSH terms have been applied to compute the citation power (i.e., the influential and productivity using the Kano diagram) with the metrics shown in Figure 4.

The next feature is to partition clusters for authors (or MeSH terms in Fig. 4) by applying SNA. The clusters can be characterized by a pattern of co-occurrences, such as the latent class model in statistics or the unsupervised training programs using artificial neuron network methods used for highlighting the type of actors (eg, author RD in this study).

The third feature is to apply the AWS to quantify co-author contributions in computing bibliometric indices (eg, h-/x-index). As such, the integral h-index, which is hard to use discriminate in RA, can be improved. The Vavryčuk’s combined weighted scheme is a special case of the AWS applied in Eq. 3. The reasons for using x-index on 2 axes in Figures 2 and 4 are (1) closely correlated to h-index; (2) newly developed in 2018.

The reasoned rationality, without considering h-index as the metric, has been discussed using examples. The x-index was thus used to discriminate against the IRA in this study.

The fourth feature uses the Kano diagrams combined with the x-index on a dashboard using Google Maps, which is rarely seen in the literature.

The fifth feature includes the Pubmed Center citations used in this study. Traditionally, many authors applied academic databases, such as the SCI (Thomson Reuters, New York, NY, the United States), Scopus (Elsevier, Amsterdam, the Netherlands), and Google Scholar to investigate the most cited articles in a specific discipline. Few were found using PubMed to retrieve the citing articles.

4.3. Limitations and suggestions

Although findings are based on the above analysis, several potential limitations may encourage further research efforts. First, this study only focused on 1 sole author, and the results cannot be generalized to other sole authors in PubMed.

Second, certain biases might exist during citation extractions because the number of citations will increase along with the date elapsed. The author’s RA and RD might differ if the time periods and the citation sources of the data are disparate.

Third, using JIF as comparisons in metrics for various scenarios of author ordering placements (top in Table 1) is a limitation because several journals in PubMed have not been indexed by the database of SCI. By contrast, certain SCI-indexed journals have not been included in PubMed. All these findings will affect the results of publications and citations obtained in this study.

Fourth, although our cluster analysis and the AWS formula in Eq. (1) are useful approaches for verifying the association of MeSH terms as the fair way for quantifying IRA, the results may be affected by the number of clusters determined before performing the SNA.

Fifth, the real corresponding author might not be the last author in an article byline. Furthermore, we exerted considerable effort into cleaning and identifying data, but typos and errors still existed, and they will affect the study results which might be biased to a certain extent.

Finally, beside DR, Tseng, many other eminent sole-authors might actually exist in academics. Their RAs and RDs are worthy of using more comprehensive analyses to create a new form of stratification based on country and/or field of research in the future.

5. Conclusions

We applied SNA to (1) examine the role of a sole-author in an authorship network, (2) investigate the RA/RD against the publication pattern in the past, and (3) understand the type of publications for a sole author in academics.

Appropriate metrics on RA are selected. The RDs are denoted by the MeSH terms and measured by the GC. The AWS used for
quantifying author credits in an article is merit to evaluate the author’s RA. The methods applied in this study can be replicated to evaluate the RA and RD for other productive authors in the future.

Acknowledgments
We thank Enago (www.enago.tw) for the English language review of this manuscript.

Author contributions
MY developed the study concept and design. WC, YT, and SC analyzed and interpreted the data. PH monitored the process of this study and helped in responding to the reviewers’ advice and comments. TWC drafted the manuscript, and all authors provided critical revisions for important intellectual content. The study was supervised by PH. All authors read and approved the final manuscript.
References

[1] American Psychological Association. Playing for “Team Science”: Tips for students. 2019/11/21 Available at https://www.apa.org/science/about/psa/2013/04/team-science

[2] Chien TW, Wang HY, Kan WC, et al. Whether article types of a scholarly journal are different in cited metrics using cluster analysis of MeSH terms to display: a bibliometric analysis. Medicine (Baltimore) 2019;98:e17631.

[3] Chien TW, Chang Y, Wang HY. Understanding the productive author who published papers in medicine using National Health Insurance Database: a systematic review and meta-analysis. Medicine (Baltimore) 2018;97:e9967.

[4] Hirsch JE. An index to quantify an individual’s scientific research output. Proc Natl Acad Sci USA 2005;102:16569–72.

[5] Egghe L, Rousseau R, Van Hooydonk G. Methods for accrediting publications to authors or countries: consequences for evaluation studies. J Am Soc Inform Sci 2000;51:145–57.

[6] Yen X, Cheng F, Wang X, et al. Top 100 cited articles on infection in rheumatoid arthritis: a bibliometric analysis. Medicine (Baltimore) 2019;98:e14523.

[7] Zou LX, Liao P, Li S, et al. Current Status of helicopter emergency medical services in China: a bibliometric analysis. Medicine (Baltimore) 2019;98:e14439.

[8] Yu G. Using meshes for MeSH term enrichment and semantic analyses. Bioinformatics 2018;34:3766–7.

[9] Kobewka DM, van Walraven C, Taljaard M, et al. The prevalence of drug-drug interactions from PubMed using MeSH terms. PLoS One 2017;12:e0173548.

[10] Prochnow T, Delgado H, Patterson MS, et al. Social network analysis in child and adolescent physical activity research: a systematic literature review. J Phys Act Health 2020;1–1.

[11] Hagens NT. Harmonic allocation of authorship credit: source-level neural networks based on OxRAM devices under a fully unsupervised training program on balance and muscle strength in older adults: a systematic review and meta-analysis. Sports Med 2017;47:2341–61.

[12] Lacroix A, Hortobágyi T, Beurskens R, et al. Effects of supervised vs unsupervised training programs on balance and muscle strength in older adults: a systematic review and meta-analysis. Sports Med 2017;47:2341–61.

[13] Kano N, Seraku N, Takahashi F, et al. Attractive Quality and Must-Be Quality. Journal of the Japanese Society for Quality Control 1984;41:39–48.

[14] Hsu WT, Chien TW, Kuo SC, et al. Whether productive authors using the national health insurance database also achieve higher individual research metrics: a bibliometric study. Medicine (Baltimore) 2020;99:e18631.

[15] Alotaibi NM, Nassiri F, Badhwa JA, et al. The most cited works in aneurysmal subarachnoid hemorrhage: a bibliometric analysis of the 100 most cited art. World Neurosurg 2016;89:587–92.

[16] Thulesius H. Assessing research impact with Google Scholar: the most cited articles in the journal 2008–2010. Scand J Prim Health Care 2011;29:193–5.