Choropleth Map Legend Design for Visualizing the Most Influential Areas on the Topic of Anesthesiology: Bibliometric Analyses

Hing-Man Wu[1], Tsair-Wei Chien[2], Willy Chou[3*]

[1]Department of physical medicine and rehabilitation, Chi Mei medical center, Tainan, Taiwan

[2]Medical Research Department, Chi-Mei Medical Center, Tainan, Taiwan

[3] Department of Recreation and Health-Care Management & Institute of recreation Industry Management, Chia Nan University of Pharmacy, Tainan, Taiwan

*Corresponding Author: Willy Chou, Department of Recreation and Health-Care Management & Institute of recreation Industry Management, Chia Nan University of Pharmacy, Tainan, Taiwan, Email: ufan0101@ms22.hinet.net

Abstract

Background: *Disparities in individual research achievements (IRA) across countries/areas are a central concern in bibliometric analyses. Lorenz curves and Gini coefficients (GC) can characterize the disparity of article citations in areas, but were rarely found in displaying on choropleth maps (CM) and interpreting the meaning of disparity related to IRA. This paper illustrates the use of cumulative frequency map legends for visualizing how the IRAs are distributed in relation to countries/areas and provinces/states on the CM.*

Objective: *To visualize the current state of research on publication outputs and citations in the field of anesthesiology to uncover the GCs across countries/areas.*

Methods: *Selecting 1,526 abstracts, author names, countries, and citations on August 20, 2018, from Pubmed Central (PMC) based on the journal of anesthesiology (JOA) in years from 2012 to 2016, we proposed a novel method that can effectively complement and efficiently improve the CM for interpreting the distribution and disparity of IRA across countries/areas using Lorenz curves and Gini coefficients (GC). We programmed Microsoft Excel VBA routines to extract data. Google Maps were used for displaying graphical representations.*

Results: *We found that (1) the dominant nations on the topic of anesthesiology in publications and citations is the US with GCs of 0.73 and 0.43, respectively; (2) Massachusetts is the dominant state in the US with a GC value of 0.45 around all states; (3) Beijing is the dominant city around China with a GC of 0.42 if Taiwan is excluded in comparison from China.*

Conclusions: *Cumulative frequency legends can be useful to complement CMs. They contain more information than those in standard CM legends that are commonly used with the method of equal intervals in classification. The Lorenz curves and GC should be combined with the use of cumulative frequency legends on the CM in the discernible future.*

Keywords: *Pubmed Central, choropleth map, Gini coefficient, Google Maps, legend design.*

1. INTRODUCTION

Anesthesiology is the medical specialty of anesthesia or anesthetics [1] and the practice of medicine that is dedicated to the relief of pain and total care of the surgical patient before, during and after surgery [2]. Anesthesia is defined by the American Society of Anesthesiologists, anesthesiology as the loss of sensation with or without the loss of consciousness [3].

Many articles have illustrated the use of choropleth map (CM) in health outcomes across areas, such as dengue outbreaks [4, 5], disease hotspots [6], and the Global Health Observatory (GHO) maps on major health topics [7], have been presented in the literature.

However, choropleth maps display the spatial properties of a data distribution but do not directly present the associated statistical distribution of the data [8]. As such, ogive-based

legends for choropleth mapping was suggested [9] along with other proposals, such as proportional symbols and gradual bubbles [10-12].

Strangely, thus far, none applied Lorenz curves [13] and Gini coefficients (GC) [14] onto the CM when we concern about the issue of data distribution of disparity across entities, such as countries/areas or provinces/cities.

The purpose of this research is to illustrate the use of legends with accumulative frequency based on quantile classification method and improve the communicative efficiency of choropleth maps of interested topics through an enhanced legend design, particularly using the Lorenz curses and GCs to explain. This design approach uses different combinations of cumulative frequency graphs in the map legend to highlight alternatively the underlying statistical distribution regarding characteristics of the areas being mapped. The approach is applied to mapping publicly available data on the topic of anesthesiology in the library of Pubmed Center (PMC).

2. METHODS

2.1. Data Sources

We programmed Microsoft Excel VBA (visual basic for applications) modules to extract abstracts and their corresponding coauthor names as well as the countries/areas of the first authors for each article on August 20, 2018, from Pubmed Central (PMC) based on the journal of anesthesiology (JOA) from 2012 to 2016. Only those abstracts published by the keyword anesthesiology [journal] and labeled with Journal Article were included. Others like those labeled with Published Erratum, Editorial or without author nation name were excluded from this study. A total of 1,526 eligible abstracts were obtained from PMC.

2.2. The Calculation of the Gini Coefficient used in this Study

In economics, the GC is a measure of statistical dispersion to represent the income or wealth distribution of a nation's residents. The calculation rule is to classify five classes with the quantile method (i.e., equal number of observations in each class) for the specific nation's residents first, and then apply the formula below to get the GC.

$$\text{Gini} = q/(q-1) \times \frac{\sum_i \sum_j |X_i - X_j|}{2 \sum_i \sum_j \bar{X}_{ij}} = q/(q-1) \times \sum_{i=1}^n |D_i|/(2 \times q * q \times \bar{X}_{ij}), \quad \text{(Eq.1)}$$

where D is the absolute difference of each pair element between classes, Xij is the value of the data element, Xij-bar is the mean of all elements across all classes, q is the number of classes, and q/(q-1) is the adjustment for the use on the number of classes to reach 1.0. Otherwise, the maximal GC reaches 0.8 if the adjustment is removed from the Eq.1[15]. The alarming level of GC for denoting inequality is set at 0.4 [15-17].

Several scenarios will be applied to the Lorenz curves and GCs, such as uniform and Chi-square distributions for understanding the feature of the GC in detail.

2.3. The Most Influential Areas in JOA Based on Countries/Areas

We applied the bibliometric x-index [18] to gain the individual research achievements (IRA). The

world-based CMs were plotted first to compare the IRA disparities of publications and citations across countries/areas.

The US and China based CMs were drawn later to see which provinces/cities or states play the most influential roles on the topic of anesthesiology. The GCs were compared onto the forth mentioned CMS in ration to the disparity of data distribution.

3. RESULTS

3.1. Task1: The Illustrations of Lorenz Curves to Explain the Pattern of Data

Four types of data distributions are shown in Figure 1. We can see the uniform type with the most equality in data distribution (i.e., GC approaching to zero). The type of Chi-square distribution on data displays the skewed pattern with a high GC near to 1.0.
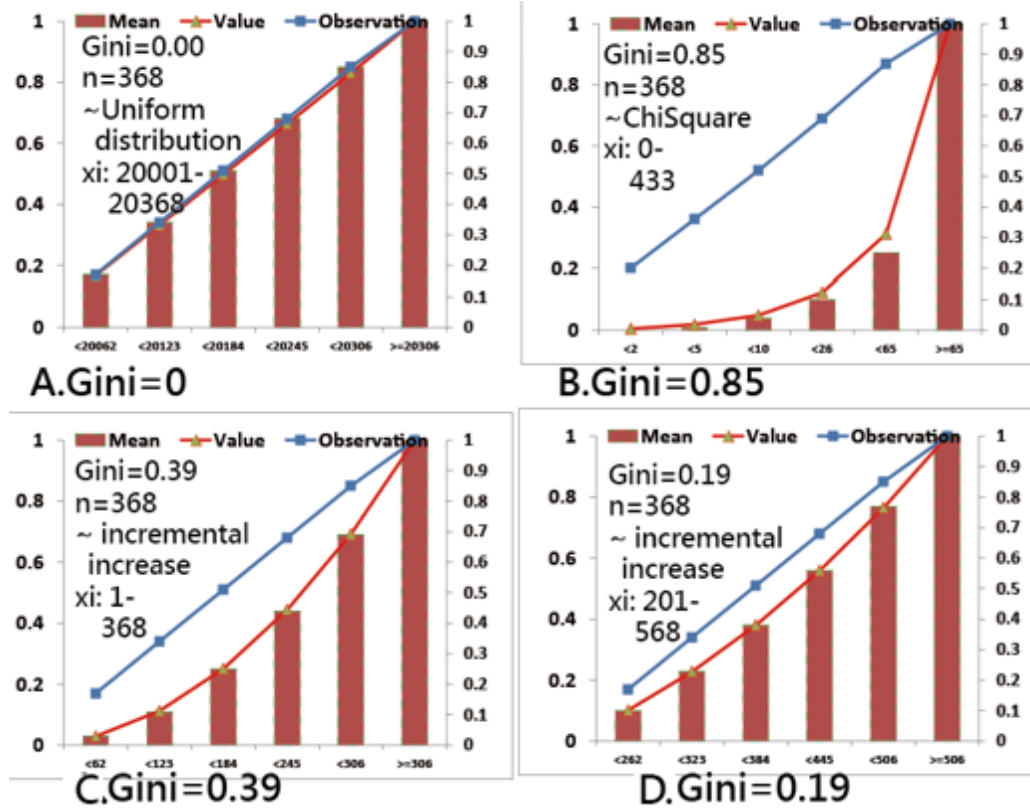
Figure1: *Illustrations of Lorenz curves on several scenarios*

3.2. Task2: The Most Productive and Cited Countries/Areas on Anesthesiology

The most productive countries/areas regarding anesthesiology are the US, France, and Canada with a GC of 0.71, see Figure 2.

The top three countries of the US, Canada, and Germany are the most cited nations with a GC of 0.43 shown in Figure 3.
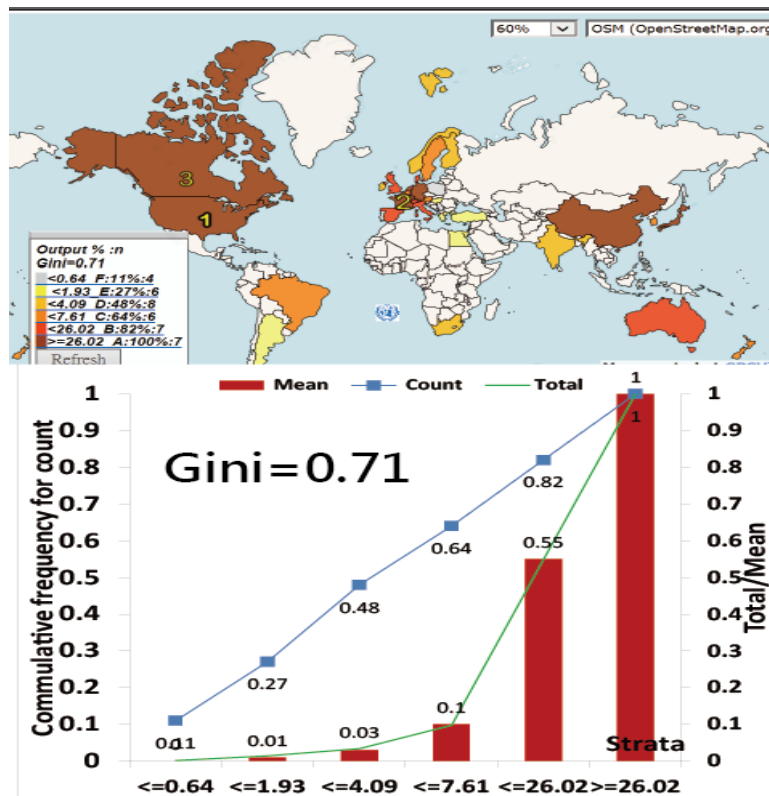

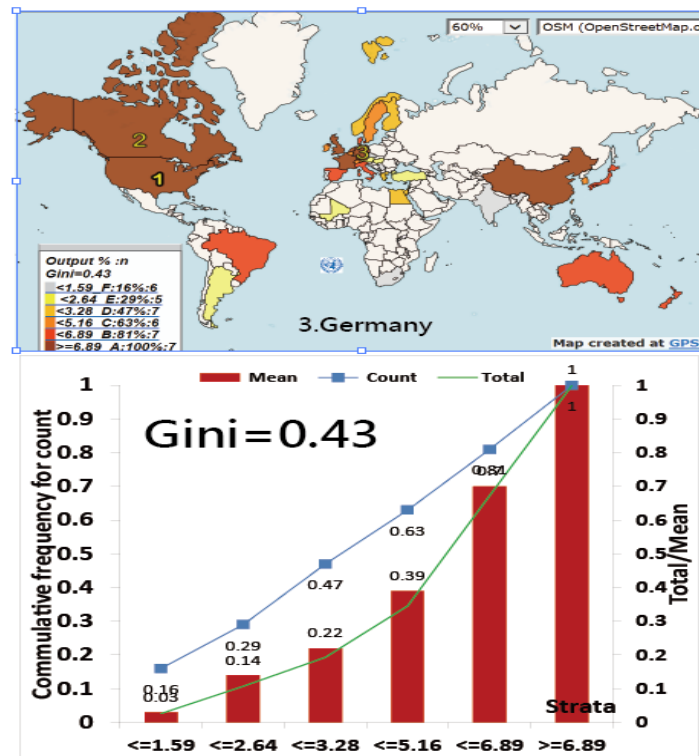**Figure2:** *The weighted publications around the world on anesthesiology*

Figure3: *The weighted x-indexes around the world on anesthesiology*

3.3. Task3: The Most Cited Areas in the Us and China

The top three states in the US are Massachusetts, California, and New York with a GC of 0.45, see Figure 4. In China, the top three are those from Taiwan, Beijing, and Shanghai with a GC of 0.42, see Figure 5.

All legends in CMs consist of the three types of information including (1) the cutting points, (2) the cumulative frequency on total values in classes, and (3) the observation counts across classes, which is unique when compared to the traditional CMS, such as dengue outbreaks [4, 5], disease hotspots [6], and the Global Health Observatory (GHO) maps on major health topics [7], only with the cutting points using the classification method of equal intervals.
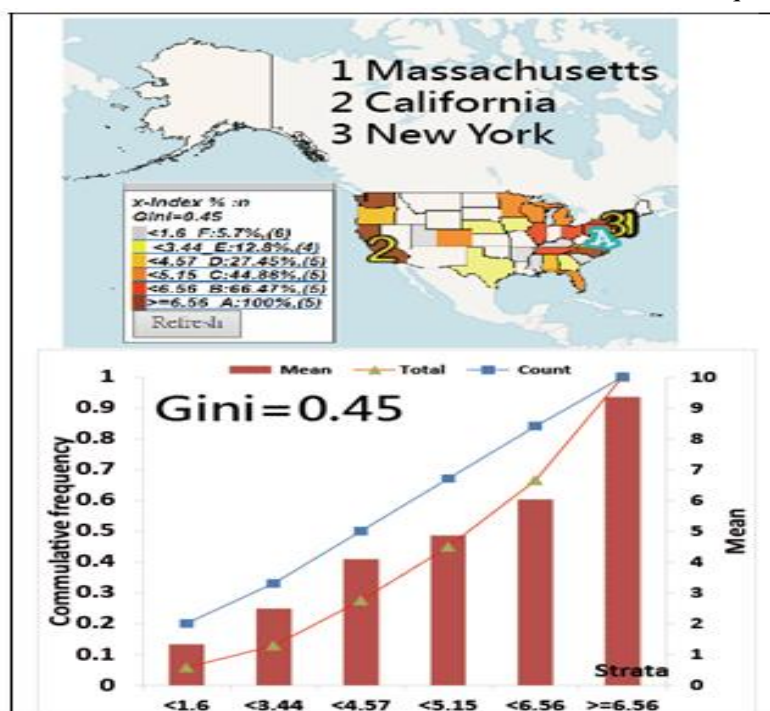

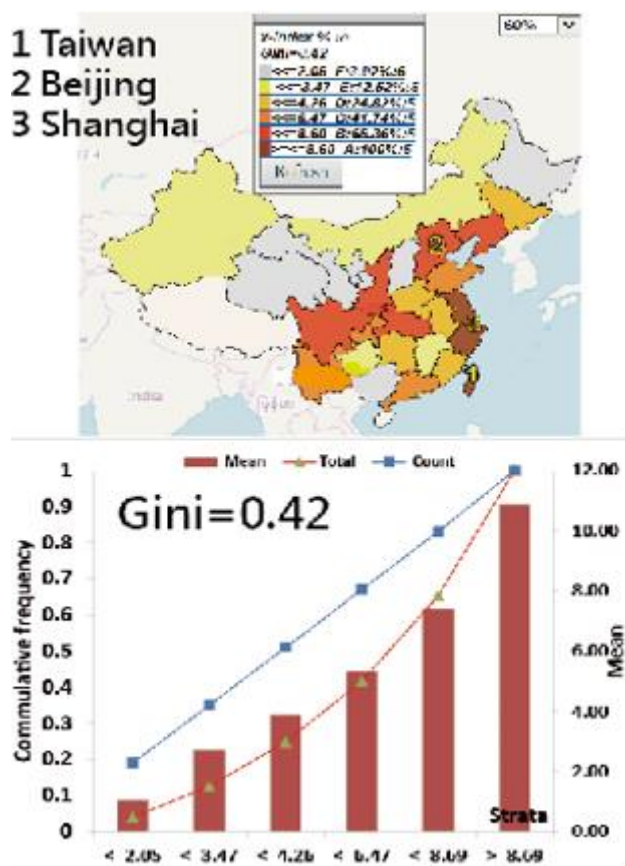**Figure4:** *Authors' x-indexes on anesthesiology by the US states*

Figure5: *Authors' x-indexes on anesthesiology by China provinces/cities*

4. DISCUSSION

We found that (1) the dominant nations on the topic of anesthesiology in publications and citations is the US with GCs of 0.73 and 0.43, respectively; (2) Massachusetts is the dominant state in the US with a GC value of 0.45 around all states; (3) Beijing is the dominant city around China with a GC of 0.42 if Taiwan is excluded in comparison from China.

Although several legends were proposed before [9-12], such as ogive-based legends, proportional symbols and gradual bubbles. No such information of Gini coefficients were provide to readers.

The equal interval method (i.e., dividing the data into equal range classes), such as ages in 0–10, 10–20, 20–30, etc., are designed by crossing the entire range. We should avoid using the equal intervals if data were skewed to one end. The outliers would likely produce empty classes for wasting classes with no observations in the specific class. We provide the quantile classification method to CMs. The feature of equal observation counts across classes allows us to not only exactly calculate the GC and plot the Lorenz curves, but also properly apply the analysis of variance (ANOVA) to differentiate the values among classes.

Several features are existed in this study including (1) the quantile classification method that allows the accumulative frequency to be equivalent to both the average value and the total values in each class.

The applications of (1) the adjusted Gini formula (i.e., Eq. 1), (2) the x-index calculation, (3) the author weight equation [3], and (4) the CMs to evaluate data disparities are unique to gain the information about the IRA among countries/areas, which are rarely seen in previously published articles before.

We used PMC citations different from those using the Scientific Citation Index (SCI; Thomson Reuters, New York, NY, the United States), Scopus (Elsevier, Amsterdam, the Netherlands), and Google Scholar [19, 20], to investigate the most-cited articles in a discipline or on a topic. The PubMed library that provides citation and reference data is a free search engine on life sciences and biomedical topics.

The reason we applied x-index in this study is the strength of the index in practice. According to the illustration in the study of Fenner and his

colleagues [18], the x-index can truly extend the feature of an author with quality and quantity IRAs in academics as mentioned above.

Although findings are based on the above analysis, there are still several potential limitations that may encourage further research efforts. First, all data were extracted from the PubMed database. There might be some biases of understanding the matched author names because of some identical names being different authors, which will affect the result by the accuracy of the indexing author names.

Second, many algorithms have been used for quantifying author weights in an article byline. We merely applied the one in the literature [3], which might be different when distinct methods were applied to the study of author credits onto articles to the results.

Third, the data extracted from *PMC* cannot be generalized to other major citation databases—such as the Scientific Citation Index and Scopus. As such, the cited authors are determined by the paper selections on Pubmed in this study only.

In conclusion, cumulative frequency legends can be useful to complement CMs. They contain more information than those in standard CM legends that are commonly used with the method of equal intervals in classification. The Lorenz curves and GC should be combined with the use of cumulative frequency legends on the CM in the discernible future.

REFERENCES

[1] Wang HY, Chien TW, Chow JC, Chou W*. Applying Clustering Coefficient to the Pattern of International Author Collaboration in Anesthesiology.ARC Journal of Anesthesiology 2017; 2(3), 18-24.

[2] American Society of Anesthesiologists. What is Anesthesiology? 2017/12/20 available at https://goo.gl/ nAxpXd

[3] Chien TW, Kan WC, Wang HY, Chou W. Visualizing Topic Burst and Citation Comparisons among Clusters of Medical Subject Headings on Anesthesiology: Bibliometric Analyses. ARC Journal of Anesthesiology. 2018; 3(4):18-24.

[4] Chen WJ. Dengue outbreaks and the geographic distribution of dengue vectors in Taiwan: A 20-year epidemiological analysis. Biomed J 2018; 41(5):283-289.

[5] Lai WT, Chen CH, Hung H, Chen RB, Shete S, Wu CC. Recognizing spatial and temporal clustering patterns of dengue outbreaks in Taiwan.BMC Infect Dis. 2018;18(1):256.

[6] Soetens L, Hahné S,Wallinga J. Dot map cartograms for detection of infectious disease outbreaks: an application to Q fever, the Netherlands and pertussis, Germany.Euro Surveill 2017;22(26): pii: 30562.

[7] WHO. Global Health Observatory Map Gallery. 2019/4/1 available at https://www.who.int/gho /map_ gallery/en/

[8] Cromley RG, Cromley EK. Choropleth map legend design for visualizing community health disparities. Int J Health Geogr. 2009; 8:52.

[9] Cromley RG, Ye Y. Ogive-based legends for choropleth mapping. Cartogr Geogr Inform Sci.2006; 33:257-268.

[10] Cromley RG, Cromley EK. Choropleth map legend design for visualizing community health disparities. Int J Health Geogr. 2009; 8:52.

[11] Barcelona Field Studies Centre. Choropleth Map with Proportional Symbols. 2019/4/4 available at https://geographyfieldwork.com/ DataPresentationMappingTechniques.htm

[12] Houle B, Holt J, Gillespie C, Freedman DS, Reyes M. Use of density-equalizing cartograms to visualize trends and disparities in state-specific prevalence of obesity: 1996-2006.Am J Public Health. 2009 Feb; 99(2):308-12.

[13] Lorenz MO. Methods of measuring the concentration of wealth. Publications of the American Statistical Association. Publications of the American Statistical Association 1905; 9(70): 209–219.

[14] Gini C. Concentration and dependency ratios (in Italian). English translation in Rivista di Politica Economica 1909; 87 (1997), 769–789.

[15] Chien TW, Chow JC, Chang Y, Chou W. Applying Gini coefficient to evaluate the author research domains associated with the ordering of author names: A bibliometric study. Medicine 2018; 97(39): e12418.

[16] Biancotti C. A polarization of inequality? The distribution of national Gini coefficients 1970–1996. J Econ Inequality 2006;4:1–32

[17] Tao CL Y, Wu X. Rawls' fairness, income distribution and alarming level of Gini coefficient. Economics Discussion Papers, No 2017-67, Kiel Institute for the World Economy.

[18] Fenner T, Harris M, Levene M, Bar-Ilan J. A novel bibliometric index with a simple geometric interpretation. PLoS One. 2018; 13(7):e0200098.

[19] Alotaibi NM, Nassiri F, Badhiwala JH, Witiw CD, Ibrahim GM, Macdonald RL, Lozano AM.The Most Cited Works in Aneurysmal Subarachnoid Hemorrhage: A Bibliometric Analysis of the 100 Most Cited Art. World Neurosurg. 2016 May;89:587-592.e6.

[20] Thulesius H.Assessing research impact with Google Scholar: the most cited articles in the journal 2008-2010.Scand J Prim Health Care. 2011 Dec;29(4):193-5