

Using the alluvial plot to visualize the network characteristics of 100 top-cited articles on attention-deficit/hyperactivity disorder (ADHD) since 2011

Bibliometric analysis

Ya-Ching Tsai, MD^a, Tsair-Wei Chien, MBA^b , Jian-Wei Wu, MD^c, Chien-Ho Lin, MD^{d,e,*}

Abstract

Background: Attention-deficit/hyperactivity disorder (ADHD) is a common neuro developmental disorder that affects children and adolescents. It is estimated that the prevalence of ADHD is 7.2% throughout the world. There have been a number of articles published in the literature related to ADHD. However, it remains unclear which countries, journals, subject categories, and articles have the greatest influence. The purpose of this study was to display influential entities in 100 top-cited ADHD-related articles (T100ADHD) on an alluvial plot and apply alluvial to better understand the network characteristics of T100ADHD across entities.

Methods: Using the PubMed and Web of Science (WoS) databases, T100ADHD data since 2011 were downloaded. The dominant entities were compared using alluvial plots based on citation analysis. Based on medical subject headings (MeSH terms) and research areas extracted from PubMed and WoS, social network analysis (SNA) was performed to classify subject categories. To examine the difference in article citations among subject categories and the predictive power of MeSH terms on article citations in T100ADHD, one-way analysis of variance and regression analysis were used.

Results: The top 3 countries (the United States, the United Kingdom, and the Netherlands) accounted for 75% of T100ADHD. The most citations per article were earned by Brazil (=415.33). The overall impact factor (IF = citations per 100) of the T100ADHD series is 188.24. The most cited article was written by Polanczyk et al from Brazil, with 772 citations since 2014. The majority of the articles were published and cited in *Biol Psychiatry* (13%; IF = 174.15). The SNA was used to categorize 6 subject areas. On the alluvial plots, T100ADHD's network characteristics were successfully displayed. There was no difference in article citations among subject categories ($F = 1.19$, $P = .320$). The most frequently occurring MeSH terms were physiopathology, diagnosis, and epidemiology. A significant correlation was observed between MeSH terms and the number of article citations ($F = 25.36$; $P < .001$).

Conclusion: Drawing the alluvial plot to display network characteristics in T100ADHD was a breakthrough. Article subject categories can be classified using MeSH terms to predict T100ADHD citations. Bibliometric analyses of 100 top-cited articles can be conducted in the future.

Abbreviations: ADHD = attention-deficit/hyperactivity disorder, CC = correlation coefficient, CD = centrality degree, IBP = impact beam plot, IF = impact factor, MeSH = medical subject heading, SNA = social network analysis, WoS = web of science.

Keywords: ADHD, alluvial plot, bibliometric, citation analysis, medical subject heading, PubMed, social network analysis, web of science

1. Introduction

An overview of the data visualization highlights the importance and generalizability of the research.^[1–3] In Sankey flow

diagrams, named after Captain Matthew Sankey, the movement or change from 1 state to another (or 1 time to another) is emphasized.^[4,5] Physics, economics, and business frequently use Sankey diagrams to analyze complex multistep processes.^[6]

The authors have no funding and conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

Supplemental Digital Content is available for this article.

^a Department of Psychiatry, Kai-Suan Psychiatric Hospital, Kaohsiung, Taiwan,

^b Department of Medical Research, Chi-Mei Medical Center, Tainan, Taiwan, ^c Department of Physical Medicine and Rehabilitation, Taipei Medical University Hospital, Taipei, Taiwan, ^d Department of Psychiatry, Chi Mei Medical Center, Tainan, Taiwan, ^e Department of childcare and education, South Tainan University of science and technology, Tainan, Taiwan.

*Correspondence: Chien-Ho Lin, Chi-Mei Medical Center, 901 Chung Hwa Road, Yung Kung Dist., Tainan 710, Taiwan (e-mail: alho@mail.chimei.org.tw).

Copyright © 2022 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: Tsai Y-C, Chien T-W, Wu J-W, Lin C-H. Using the alluvial plot to visualize the network characteristics of 100 top-cited articles on attention-deficit/hyperactivity disorder (ADHD) since 2011: Bibliometric analysis. *Medicine* 2022;101:37(e30545).

Received: 26 March 2022 / Received in final form: 9 August 2022 / Accepted: 10 August 2022

<http://dx.doi.org/10.1097/MD.00000000000030545>

Key Points:

- An alluvial plot was drawn to illustrate the network characteristics of T100ADHD, which is novel and contemporary.
- Bibliometric studies can benefit from the method of drawing the alluvial that is demonstrated in detail with an MP4 video.
- This study introduces an additional visualization, the impact beam plot (IBP). Online IBP was demonstrated to be effective and worthy of future studies to highlight the most influential articles with a quick glance at a picture.

However, Sankey diagrams have a limited presence in health-care literature since the creation of a Sankey may require special software or advanced programming skills.^[2,7] We are interested in using these methods (e.g., programming skills combined with Sankey software) to improve the visualization of data in bibliometric analyses.^[8,9] The network characteristics of 100 top-cited articles on attention-deficit/hyperactivity disorder (T100ADHD) were illustrated using a visualization technique (such as Sankey/alluvial diagrams, stacked bar charts, and parallel sets).^[10]

1.1. ADHD with a prevalence of 7.2%

Attention-deficit/hyperactivity disorder (ADHD) is a behavioral and neurodevelopmental disorder characterized by inattention, hyperactivity, and impulsivity in children or early adolescents.^[11–13] The prevalence of ADHD worldwide is estimated at 7.2%,^[14] affecting approximately 5% of children and 2.5% of adults.^[15,16] A survey used the combination of parent and self-reports from rating scales (using an or rule and a norm-based threshold of 4 symptoms threshold) and reported that 60% of the ADHD group demonstrated symptom persistence as adults, and 41% met both symptom and impairment criteria.^[17] Males are more likely to suffer from ADHD than females, with a ratio of 3:1 to 5:1.^[18,19]

A multifactorial disorder of ADHD with a genetic component can be recognized from its complex clinical symptomatology.^[20] ADHD symptoms vary throughout life, with younger children displaying more inattention, hyperactivity, and impulsivity than adolescents and adults.^[21] Through academic research, numerous technical and methodological advances have been made available for clinicians and researchers to better understand ADHD and its etiology. Although a majority of published articles have investigated the pathophysiology of ADHD in the literature, the characteristics of ADHD remain unclear for complex psychiatric disorders,^[22] and a broad overview of ADHD on topical entities in development and evolution over the past years is difficult.

1.2. Bibliographical studies on ADHD

Bibliometric analysis allows documents to be grouped by theme (e.g., ADHD) on the basis of a topical entity (e.g., author or affiliated country) and a feature (e.g., citations and publications). However, only 2 articles^[20,23] were directly related to ADHD in PubMed when searching for the keywords ADHD and (bibliometric medical searching heading [MeSH Major Topic]). A common drawback of traditional bibliometric analyses^[23–28] is the accumulation of too much tedious information (e.g., 15 tables in this study^[4]) that makes it difficult to focus on the major points. The readers will be interested in taking a quick look at a few pictures, since nearly everyone enjoys reading photographs.

It is said that a look is worth a thousand words and quite a few numbers as well.^[29] Until recently, such illustrations – Sankey diagrams and alluvial plots – could be constructed by piecing them together in a drawing program.^[30] The motivation of this study is to provide a method of visualizing the network characteristics of T100ADHD.

1.3. Alluvial versus Sankey diagrams

1.3.1. Brief overview of the Sankey flow diagram. Flow diagrams depict paths between related events (i.e., states, positions, or steps).^[31] In a Sankey diagram, nodes represent entities, and arcs represent relationships. During transitions, each arc flows from its source node to its target node(s),^[2] in contrast to a stacked bar chart that focuses solely on proportional changes in values across nodes (or entities).

To construct the Sankey network, we follow a series of steps: occasion levels (i.e., time points) and values on axes x and y ; nodes with weights (e.g., $A \rightarrow B$ 30 as expressed in social network analysis (SNA)^[32]). The size of each node and the width of each arc represent the number of objects/members (or weights as defined in bibliometrics^[33]), indicating the magnitude of the flow.^[5] In a stacked bar chart, the nodes are only stacked vertically and organized by steps on the x -axis in contrast to the alluvial plot,^[34] which illustrates how a population of facts (e.g., the number of articles or citations) is allocated across categorical dimensions on the x -axis; left/right position is unimportant; and dimensions (e.g., years, countries of origin, research categories, document types, publishing journals, etc. in bibliometrics) may be arranged in any order.

1.3.2. Harder to draw the alluvial plot in bibliometrics. Using the same logic of the Sankey diagram, alluvial plots^[35] illustrate how the same set of items (called entities in this study) regroup according to a variety of dimensions (e.g., years, countries of origin, research categories, document types, and publishing journals in bibliometrics). As a result, the alluvial plot is more difficult to draw for bibliometric analysis than the Sankey plot because the weights on arcs and nodes in SNA equal the number of publications (or citations),^[33] and flows between dimensions on the x -axis backward extracted from the SNA^[32] instead of the step-by-step process in a forward approach; and Sankey places nodes more freely than the alluvial, which requires their nodes to be aligned and vertically matched to the respective dimension on the x -axis and cannot be randomly placed.

Alluvial plots are types of Sankey diagrams that illustrate how items regroup on a visual representation.^[36] The 2 terms are thus often used interchangeably in practice, but in this study, we refer to the alluvial as dimensions rather than paths between related events (e.g., steps).

1.4. Study aims

In this study, we will demonstrate how to quickly draw the alluvial used for bibliometric analysis and use the alluvial to obtain a better understanding of network characteristics among entities of interest in T100ADHD.

2. Methods**2.1. Data source**

The data were arranged in 2 steps. First, the authors searched the PubMed database using the keywords (“2011” [Date - Publication]: “2021” [Date - Publication]) and (ADHD) AND (MeSH Major Topic) as of March 12, 2022 and matched it with the number of citations in Web of Science (WoS). The relevant metadata (e.g., years, countries of origin, research categories,

document types, publishing journals, [MeSH] terms) were collected in T100ADHD.^[37]

Second, based on the article metadata, 3 components were included: descriptive statistics in traditional bibliometrics, visualizations using the alluvial to display all relevant entities and their associations, and inferring statistics regarding the prediction of article citations using MeSH weights.

This study does not require ethical approval since all data were obtained from a publicly available database.

2.2. Weights equal to article numbers and citations in SNA

2.2.1. All weights equal to article number. Since all elements of entities (such as years, countries of origin, research categories, document types, publishing journals, etc) in articles have equal weights, it was determined that the number of connections for a specific element in an article could be computed (e.g., $CW_i = \frac{1}{L}$, and $CW_i = \text{citations} \times \frac{1}{L}$ for publications and citations, respectively, where L is the number of elements in an article, and CW_i is the weighted counts or citations for entity (i). As shown in Equation 1, the equal weight (W_i) is equal to 1. The weighted centrality degree (CD) was calculated using Equation 1. Each element defined as a node in SNA earns a WCD. The CD for a given element is determined by adding the weights of all articles from 1 to n, where n denotes the sample size. Therefore, a particular adjustment is required to make all WCDs in Equation 2 equal the number of articles via Equations 1 to 3.^[31]

$$W_i = \frac{1}{L}, (1)$$

$$\begin{aligned} \text{WCD in an article} &= \left(\sum_{i=1}^{L-1} \sum_{j=i+1}^L (W_i + W_j) \right) \div (L - 1) \\ &= 2 \times W_i \times \frac{L \times (L-1)}{2} \div (L - 1) \\ &= 2 \times \frac{1}{L} \times \frac{L \times (L-1)}{2} \div (L - 1) = 1, (2) \end{aligned}$$

where L is the number of cooccurring elements of entities (e.g., countries or MeSH terms). In an article byline, W_i represents the equal weight given to each entity; see Equation 1. Therefore, WCD is the weighted CD. According to Equation 2, W_i equals W_j . All entities in an article have a WCD of 1.0. The total WCD is equal to the number of articles, as opposed to the traditional SNA, which ignores the WCD equal to the number of articles.

For instance, there are 2 MeSH terms A and B in an article. The $WCD = 2 \times \frac{1}{2} \times \frac{2 \times (2-1)}{2} \div (2 - 1) = 1$. Similarly, WCD equals 1 if L is >2. The WCD for a specific MeSH term (or other entity) k is the summation of W_i in all articles in Equation 3.

$$WCD_k = \sum_{j=1}^n W_j, (3)$$

Herein, n equals 100 in this study. A bubble in SNA is sized by the WCD.

2.2.2. All citation weights equal to article number multiplied by citations. If article citations (denoted by IF) were taken into account, the weights of a specific entity (e.g., WCD in Eq. 2) would be combined with IF to compute the $IFWCD_k$ via Equations 4 and 5.

$$\begin{aligned} IFWCD \text{ in an article} &= IF \times \left(\sum_{i=1}^{L-1} \sum_{L=i+1}^L (W_i + W_L) \right) \div (L - 1) \\ &= IF \times 2 \times W_i \times \frac{L \times (L-1)}{2} \div (L - 1) \\ &= IF \times 2 \times \frac{1}{L} \times \frac{L \times (L-1)}{2} \div (L - 1) = IF, (4) \end{aligned}$$

$$IFWCD_k = \left(\sum_{j=1}^n IFWCD_j \right) \div WCD_k, (5)$$

$$\text{Citations} = \sum_{k=1}^n IFWCD_k, (6)$$

where $IFWCD_k$ is the mean IF of entity k. All citations in articles are composed of individual $IFWCD_k$ in Equation 6. Accordingly, the bubble can be sized by the $IFWCD_k$ for each entity.

2.2.3. Examples illustrated in various scenarios. Supplemental Digital Contents 1, <http://links.lww.com/MD/H363> and 2, <http://links.lww.com/MD/H364> provide detailed information regarding the weight settings for each cell in the MSeXcel worksheet. When the trend is known for each entity in the row, the weights for each element in the cells are different due to different observed counts (=weights set in the cells). With the Sankey software shown in Section 2.4, a trend-type alluvial plot can be obtained using the scheme expressed in Supplemental Digital Content 2, <http://links.lww.com/MD/H364>.

2.3. Taking publications and citations into account using the ht-index

As a bibliometric indicator, the hT index^[38,39] is calculated using Equations 7 to 9. A vector of citations (e.g.,^[1,2,4,6] in descending order in 5 articles) is used to calculate the hT-index = 4.03. The hT index in the most cited paper with n1 citations is calculated by summing all weights from the starting 1 to the following of $\frac{1}{2i-1}$ via Equation 7, where i is from 1 to n1.

$$h_{T(top\ 1)} = \sum_{i=1}^{n_1} \frac{1}{2i-1}, (7)$$

According to Equation 8, the starting weight for the top-cited paper is assigned by 1.0. This results in a sum of 2.13 for 10 citations, 3.28 for 100 citations, 4.44 for 1000 citations, and 5.59 for 10,000 citations.^[40]

$$\text{Weight}_{(top\ j)} = \frac{1}{2j-1}, n_j \leq j (8)$$

$$h_{T(j)} = \frac{j}{2j-1} + \sum_{i=1}^{n_1} \frac{1}{2i-1}, n_j > j (9)$$

There are 2 characteristics of the hT index: the h-core^[41] articles are identical to those in the hT core, and the contribution of the h-core is unchanged in the hT core Durfee square.^[38,39]

The hT score for any single paper ranked j in the list (with n_j citations) is denoted as $hT(j)$ in Equation 9 if the author has N papers with associated citations $n_1, n_2, n_3, \dots, n_N$ (ranked in descending order). As a result, the hT-index for all the cited papers is calculated by summing all the weights (allocated in Eqs. 7–9) through equation 10.

$$h_T = \sum_{i=1}^N h_{T(i)}, (10)$$

2.4. How to draw the alluvial plot

2.4.1. A single example. We assume that 3 countries, the US, China, and the UK, with articles published in 2 years, 2021 and 2022, are linked by the terms (2021 7 US), (2021 6 China), (2022 2 US), (2022 9 UK), and (2022 4 China). There are some connections

that are stronger than others. Figure 1 shows that 2022 has a thin connection with the US and a much thicker connection with the UK.

2.4.2. A multilevel alluvial plot. A multilevel alluvial plot with 4 dimensions, composed of entities (i.e., countries), is illustrated in Panel B of Figure 1, in contrast to Panel C of Figure 1, which shows various entities across 4 dimensions.

2.4.2.1. Publication-based alluvia diagram. Based on article numbers, an alluvial plot was drawn as follows (see MP4 video at the link^[42]):

- (i) In Alluvia, various entities are used to display dimensions on the x-axis.
- (ii) Assign a weight (e.g., 1 and ci for publication/citation-based the alluvial) to each T100ADHD.

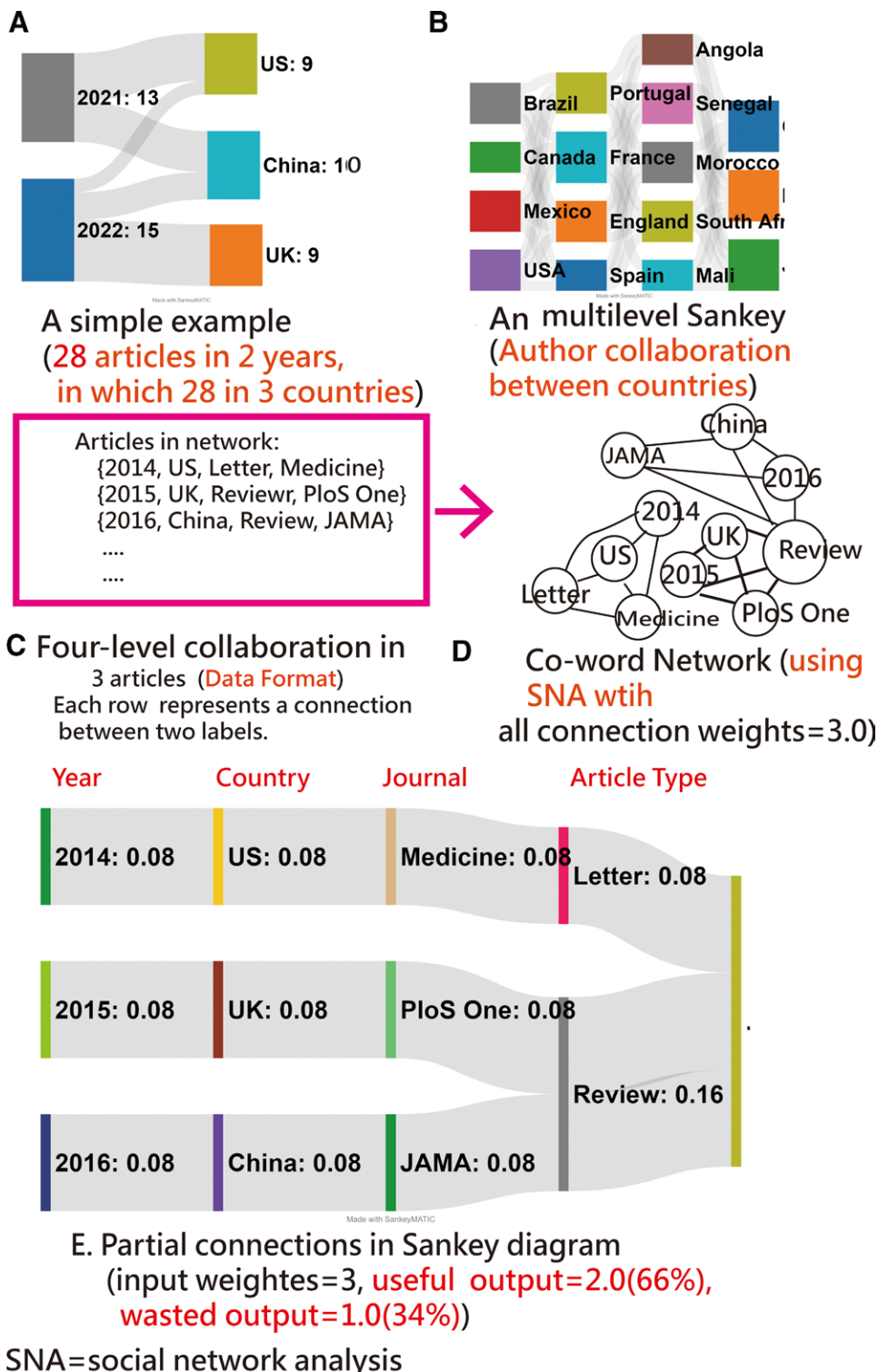


Figure 1. How to draw the alluvial plot (From a simple relationship between entities to an multiply relationship with author collaborations, the relationship is assessed using the SNA, and the centrality coefficients between entities are computed. The values of nodes and edges are consistently in the alluvial plot via an algorithm to fit the format and requirement from the software).

(iii) Calculate weights for each element among entities using the SNA (e.g., the US to China with 0.3 and the US to Medicine with 0.2).

(iv) Convert the weights of arcs into composite scores of nodes so that all weights and composite scores are equal to 100 (i.e., 100 articles in T100ADHD if publication-based the alluvial is applied).

(v) Make the codes conform to the format of the software.^[43]

(vi) The codes generated in Step (v) should be pasted into the software^[43] to draw the alluvial.

2.4.2.2. Citation-based alluvial plot. As in Steps in publication-based alluvial, the article weight = 1 to T100ADHD is replaced with their respective citations (see MP4 video at the link^[42]), and the citation-based alluvial can be produced.

2.4.2.3. The ht-based alluvial plot. Each hT for each element in entities was calculated using Equations 4 to 6. By using Step(iv) in Section 2.4.2.1, the hTs were backward transformed into arcs between elements within entities. An hT-based alluvial can thus be made for T100ADHD.

2.5. Data presentations using tables and figures

2.5.1. Overall descriptions of T100ADHD. An overview of countries in the T100ADHD was made in a table. Journals with the number of articles and citations are represented on a scatter plot. Based on normalized citations for each article, the T100ADHDs have been dotted on the impact beam plot (IBP)^[44,45] to illustrate the article impact from 0 to 100 by year. According to the online technique,^[46] the overall hT-index and the h-index^[41] were compared with the median score on the IBP.

2.5.2. Visualizations using the alluvial plots. SNA was used to cluster article subject categories (based on research areas in WoS and MeSH terms in PubMed). As described in Section 2.4.2, 3 alluvial plots were drawn.

2.5.3. Comparison and prediction in article citations. To determine whether article citations differed between subject categories, a one-way analysis of variance was conducted. Article citations were converted into $\log(\text{citation})$ because hTs are not normally distributed. We computed the impacts (IFs) of MeSH terms using Equation 6. The WCDs derived from MeSH weights were used to predict the original article citations. We used the correlation coefficient (CC) and the regression coefficient R_{sq} ($= CC \times CC$ denoted by variance explained by the model) as measures of the relationship between the WCD of MeSH terms and the number of article citations.^[19,31]

2.6. Statistics and tools

To present the research results in this study, a dashboard was developed. Except for the alluvial plots, all graphs are presented using author-made modules in Excel (Microsoft Corp.). HTML pages based on Google Maps were created.

The CC t value was denoted by the formula ($= CC \times \sqrt{\frac{n-2}{1-CC \times CC}}$). We set the significance level at Type I error (0.05). Using MedCalc statistical software, version 9.5.0.0 (MedCalc, New York, NY), a prediction equation was developed. We set the significance level at Type I error (0.05).

We used the scatter plot to illustrate the relationship between {article citations, MeSH weights} and {publications and citations of journals} in T100ADHD. It is possible to link all relevant information on the entity attributes to dashboards on Google Maps. BibExce software is used to create the dashboard for Google Maps.^[47] In Supplemental Digital Content 1, <http://>

links.lww.com/MD/H363, the MSeExcel module used to draw the alluvial plot can be found.

3. Results

3.1. Overall descriptions of T100ADHD

The top 3 countries (the United States, the United Kingdom, and the Netherlands) accounted for 75% of all T100ADHD cases. Brazil was the country with the highest number of citations per article (=415.33). As shown in Table 1, the overall impact factor (IF = citations/100) is 188.24.

IBP for T100ADHD indicates that the top-cited article authored by Polanczyk et al from Brazil in 2014 has 772 citations in WoS.^[48] As soon as the dot on the far right side of Figure 2 is clicked (or linked to the link^[37]), an abstract with 375 citations (up to March 23, 2022) appears on PubMed, which is significantly lower than that on WoS (=772). T100ADHD has an h-index of 105 and an hT-index of 126.5.

In the future, these metrics may be compared to those of other 100 top-cited articles. There are only 11 articles (red dots in Fig. 2) that are clinical research, which is much lower than the number of nonclinical articles (=89) in T100ADHD. With the IBP, we can quickly see the T100ADHD, which has never been seen in bibliometric studies regarding studies on the T100 cited articles. According to the scatter plot in Figure 3, the majority of articles were published and cited in Biol Psychiatry (13%; IF = 174.15).

3.2. Visualizations using the alluvial plots

The visualization in Figure 4 shows that 7 subject categories were classified using the SNA (Fig. 4), including Neurosciences & Neurology (24%), Psychiatry (12%), Psychology (18%), genomics (17%), Pediatrics (19%), Epidemiology (8%), and Radiology, Nuclear Medicine & Medical Imaging (2%).

As shown in Figures 5 to 7, their publications, citations, and hT-indices are presented based on 5 dimensions, respectively. There are larger blocks with greater weights, such as journal articles, the US, Neurosciences & Neurology, and Biol Psychiatry.

Figures 5 and 7 illustrate the block areas for articles, citations, and hT-indices, respectively, in useful presentations. In general, 15 tables or figures are requested to display information on 5 dimensions (i.e., years, countries, research categories, document types, and journals in rankings) in traditional bibliometrics.

Alluvial plots only display a partial relationship between dimensions (e.g., including years, journals, subject categories, countries of origin, excluding articles) rather than the entire relationship between elements in all entities.

3.3. Comparison and prediction in article citations

Among the subject categories of T100ADHD, there was no difference in the number of article citations ($F = 1.19$, $P = .320$). Based on the scheme of equal weight in articles, MeSH terms are evident in the predictive power of article citation number ($CC = 0.45$, $t = 5.04$; $n = 100$, $R_{sq} = 0.21$, i.e., the variance explained by the model) after being associated with article citations based on the MeSH weights. As a result, the regression equation is as follows: article citation (y) = $y = -0.06860 + 1.1443 \times \text{weight (x) of MeSH terms}$. According to Figure 8, the slope coefficient was statistically significant ($F = 25.36$; $p.001$).

3.4. Online dashboards shown on google maps

Upon clicking the QR code (e.g., Figs. 2–4 and 8), all dashboards in graphs will appear/readers are advised to examine the

Table 1

Distribution of T100ADHD for countries over years (n = 100).

Country	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	n	ci	IF
US	8	9	3	8	5	4	2	1	2		42	6880	163.81
UK	7	3	4	3		4	1	2			24	5257	219.04
Netherlands	2	1			2	1	1	1	1		9	1331	147.89
Denmark					3				1		4	1147	286.75
Brazil				1	1	1					3	1246	415.33
Canada	1		1			1					3	498	166
China			1					1		1	3	316	105.33
Sweden	1							2			3	334	111.33
Switzerland		1	1	1							3	366	122
Australia	1			1							2	264	132
Germany				2							2	370	185
Finland						1					1	113	113
France	1										1	159	159
n	21	14	10	16	11	12	4	7	4	1	100	18825	188.25

Note. ci = citations; IF = impact factor = ci/n, n = publications, T100ADHD = 100 top-cited articles on attention-deficit/hyperactivity disorder.

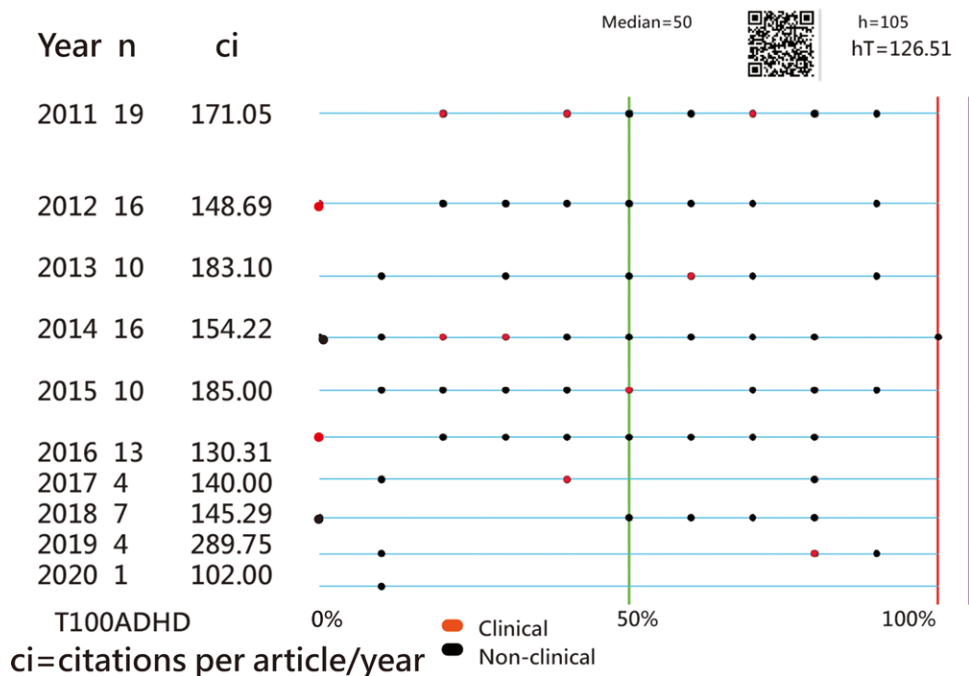


Figure 2. IBP used to display T100ADHD in this study (An article is used to denote the dot. When the dot is clicked, the article appears on the website. In IBP, the citations are arranged from the lower to the right side, with a greater number of citations on the horizontal axis). IBP = impact beam plot, T100ADHD = 100 top-cited articles on attention-deficit/hyperactivity disorder.

details of each entity. For instance, when the bubble of interest is clicked in Figures 2 and 8, the article is immediately displayed on the website. When the bubble in Figures 3 and 4 is clicked, the information associated with the bubble will appear on the dashboard.

4. Discussion

We observed 10 major results: the top 3 countries (the US, the UK, and the Netherlands) occupied 75% of T100ADHD; Brazil earned the most citations per article (=415.33); the overall T100ADHD IF (= citations/100) was 188.24; the most cited article authored by Polanczyk et al In 2014, Brazil had 772 citations in WoS; most articles were published and cited in Biol Psychiatry (13%; IF = 174.15); 6 subject categories were used.; network characteristics of T100ADHD were successfully displayed on alluvial; no difference was found in article citations among subject categories ($F = 1.19, P = .320$); physiopathology,

diagnosis, and epidemiology were the most frequently occurring MeSH terms; and MeSH terms were evident in the prediction power on the number of article citations ($F = 25.36; P < .001$).

4.1. Additional information

In this paper, 2 major goals were achieved: we demonstrated how to draw the alluvial and applied the alluvial to gain a better understanding of the network characteristics of the T100ADHD.

Two studies^[8,9] applied the alluvial to the analysis of bibliometric data. Graphs based on the alluvial are appropriate due to the categorical dimensions instead of steps (or years) on the x-axis, as stated in definitions of the alluvial plots.^[10]

Although Sankey diagrams are more widely known, alluvial plots are generally easier to generate.^[5] It is only true if the data are as simple as the example in Panel C of Figure 1, and the alluvial is drawn without taking into account any weights (e.g., citations or hT-indices) using the software.^[49] Due to this, the alluvial

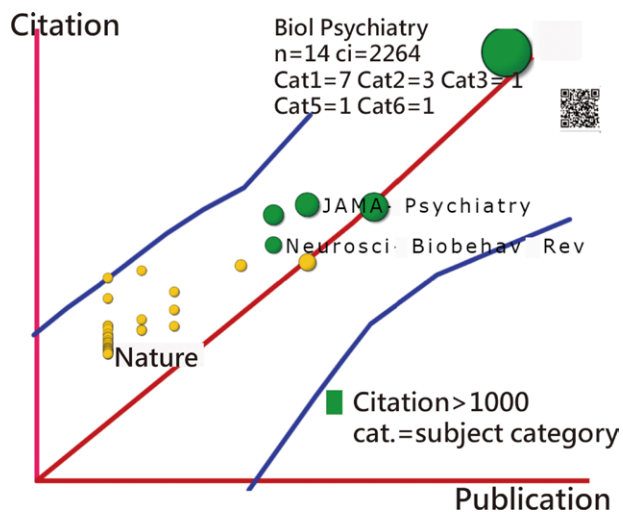


Figure 3. Distribution of journals in T100ADHD shown on the scatter plot (On the diagram, publications and citations for journals are dispersed). T100ADHD = 100 top-cited articles on attention-deficit/hyperactivity disorder.

is more difficult to draw than the Sankey, especially in 3 circumstances: the weights are calculated using SNA and allocated proportionally to nodes and arcs; the flows between dimensions are extracted backward from the SNA^[32] rather than forward by the step-by-step process employed in the alluvial; the nodes in the alluvial plot may be placed more freely than in an alluvial plot, which requires their nodes to be aligned and cannot be placed randomly. As far as we are aware, no software has been developed to take these factors into account when drawing the alluvial. An MP4 video^[42] and an MExcel module are provided in Supplemental Digital Content 1, <http://links.lww.com/MD/H363> so that readers could draw the alluvial on their own.

The results of the 3 alluvial plots in Figures 5 to 7 were similar. Based on the principle of bibliometric indicators (i.e.,

publications and citations or others), the hT index () can be used to measure research achievements since both publications and citations are considered. To make the article as concise as possible, only 1 diagram (e.g., the hT-based alluvial plot that combines publications and citations in Fig. 7) should be provided to the authors instead of the 2 (i.e., Figs. 5 and 6).

Additionally, the reasons for demonstrating the hT-index in this study are the hT-index has an identical h-core with the h-index,^[41] there is a strong association with the h-index,^[36] and all publications and citations are taken into account^[38,39] to overcome the disadvantages of many bibliometric indices.

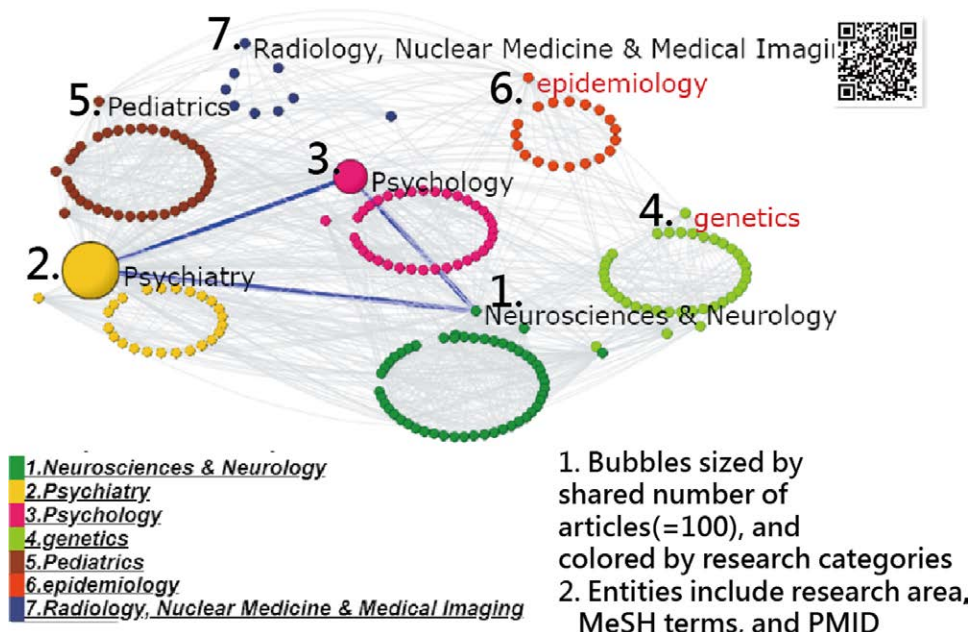
There are numerous articles on ADHD in the literature. Nevertheless, none of the studies used the WoS (or Pubmed) database to explore influential entities with alluvial plots, as we did in Figures 5 to 7. In contrast to previous findings,^[20] the top 3 countries (the US, UK, and Netherlands) differed in terms of article sources from databases (e.g., PubMed vs WoS). Only 24 articles^[20] were included in the T100ADHD based on the PubMed database.

In contrast with the 100 articles listed in a bibliometric study,^[28] the dashboard-type IBP, shown in Figure 2, is unique and modern. It presents the T100ADHD in a single view and provides more context than a single metric alone, such as a citation metric (or the h-index^[41]) in bibliometrics. Bibliometric analysis using the IBP can be a frontier in this area.

4.2. Three most-cited articles

“ADHD prevalence estimates across 3 decades: an updated systematic review and meta-regression analysis” authored by Polanczyk et al from Brazil in 2014 received 772 citations in WoS^[48] and is classified as [6.epidemiology] in this study. The authors identified 154 original studies and found that methodological procedures investigated were significantly associated with heterogeneity of studies; geographical location and year of study were not associated with variability in ADHD prevalence estimates; and variability in ADHD prevalence estimates

Classification of Research Category



1. Bubbles sized by shared number of articles(=100), and colored by research categories
2. Entities include research area, MeSH terms, and PMID

Figure 4. Classification of subject categories using the SNA (the MeSH terms were used to cluster the research categories using the SNA and 7 categories are found). MeSH = medical subject heading, SNA = social network analysis.

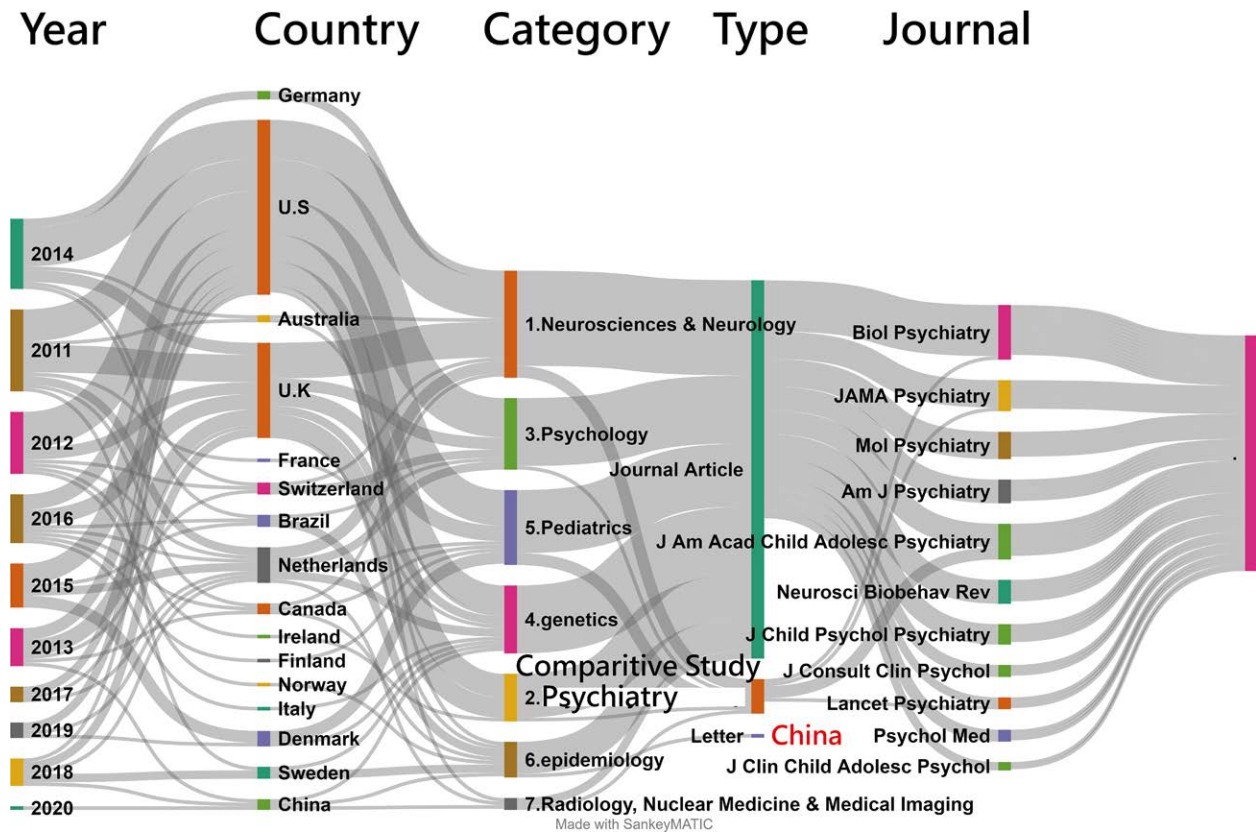


Figure 5. The publication-based alluvial plot (the publications for the most productive entities are shown on the alluvial plot).

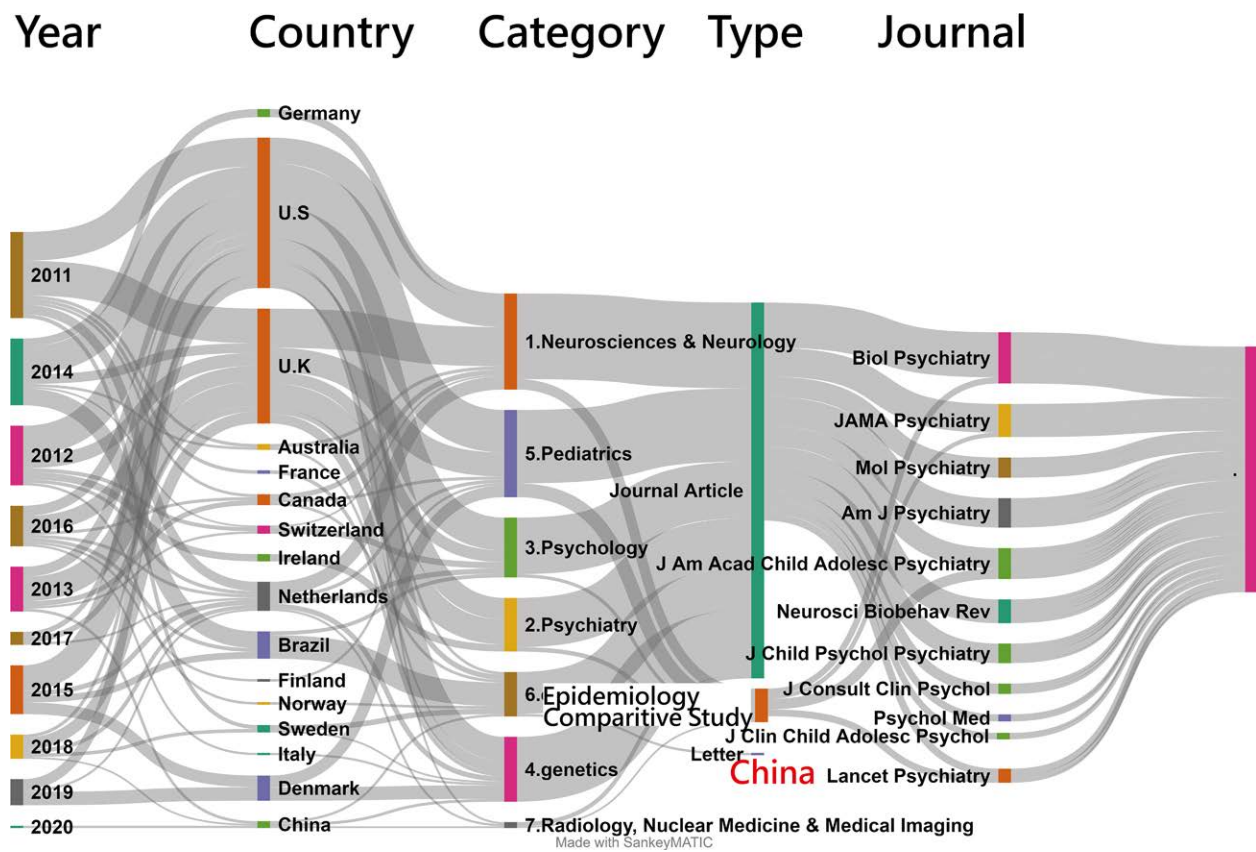


Figure 6. Citations-based alluvial plot (the citations for the most influential entities are shown on the alluvial plot).

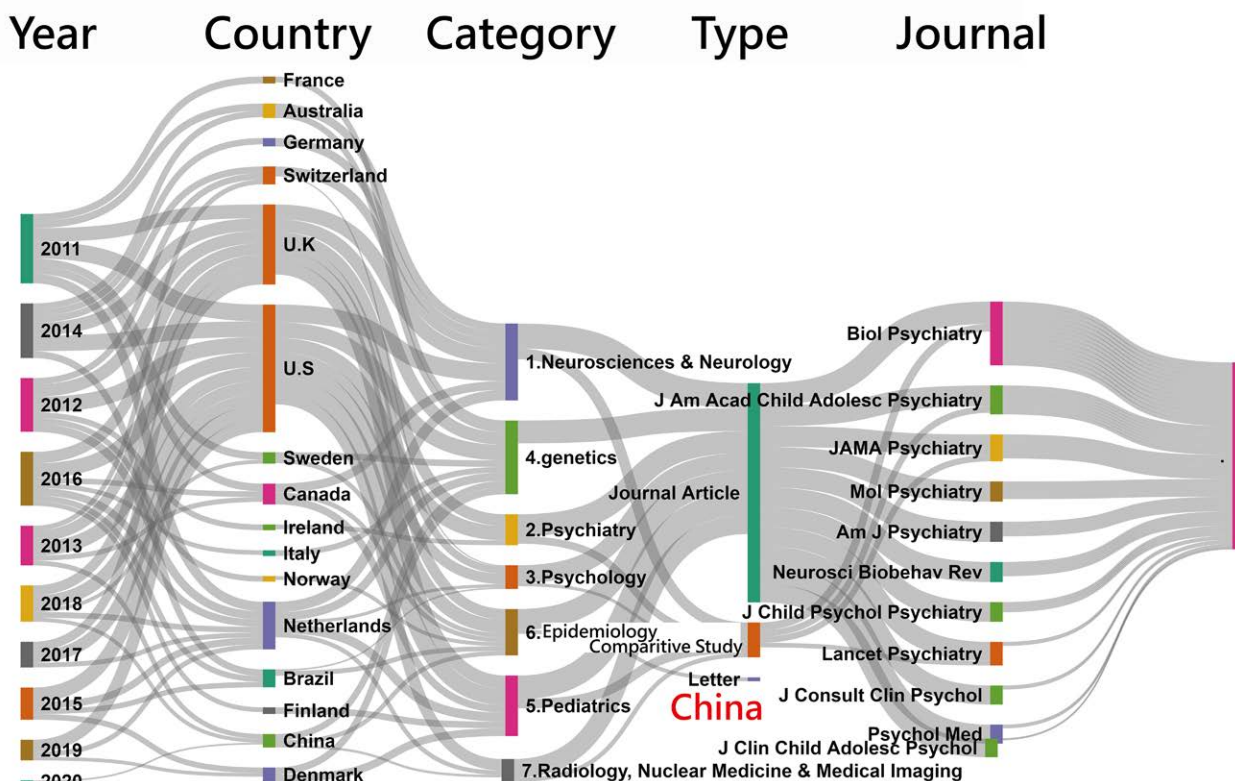


Figure 7. The hT-based alluvial plot (the hT-index taking both publications and citations into account for entities with high research achievements are shown on the alluvial plot).

is largely explained by methodological characteristics of the studies.

The second-ranked article was authored by 82 scholars and was titled “Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder,” which was published in 2019^[15] and classified as [4. Genetics] in this study. It was cited 614 times. An analysis of 3 replication studies supports their findings and highlights study-specific differences in genetic overlap with educational attainment. The

authors found important new information about ADHD’s biology. Due to the rapid advancement of gene research, this article has also been cited in many gene- and behavior-related studies.

The third-ranked article was authored by Faraone et al and titled “Attention-deficit/hyperactivity disorder,” which was published in 2015^[16] and classified as [5. Pediatrics] in this study. It was cited 579 times. There is no curative treatment for ADHD, but evidence-based treatments can significantly reduce its symptoms and associated impairments. Medications are effective and well tolerated, and nonpharmacological approaches are also beneficial. There is ongoing clinical and neurobiological research that has the potential to advance diagnostic and therapeutic approaches to ADHD, according to the authors.^[16] This study establishes a benchmark for future ADHD research.

Article citation

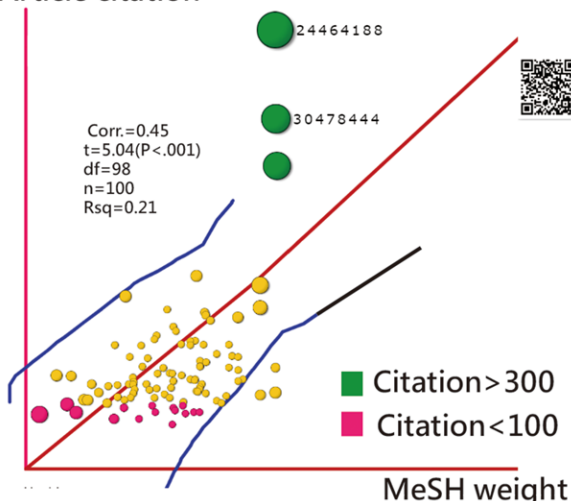


Figure 8. MeSH terms were used to predict article citations shown on the scatter plot, and significant associations were found between the weights of MeSH terms and article citations. MeSH = medical subject heading.

4.3. Implications and changes

The study has several distinctive features. First, the hT-index with decimal places offers greater discrimination power than the original h-index for identifying the research achievements and rankings of a particular group.^[50] We therefore suggested using the hT index to evaluate research achievements on individuals or at research institutions.

Second, alluvial was proposed through bibliometrics as a way to highlight the few most important entities.

Third, IBPs^[45] provide authors with a brand-new representation of every academic article, illustrating research achievements by the hT-index rather than the median percentile only shown to authors of WoS entries.^[51,52]

The fourth point is that although Sankey diagrams are better known, alluvial plots are generally considerably easier to generate.^[5] This is true only if the data are as simple as the example in Panel C of Figure 1, where the alluvial is drawn without taking into account the weights (e.g., citations and/or hT indices).

In addition, the classification of subject categories using SNA is objective and unique in comparison with previous studies using manual methods^[8] and document types only in PubMed.^[9] However, the evidence indicates that the classification is meaningful, feasible and worthy of recommendation for future studies in relevant fields ($F = 1.19, P = .320$).

Although the hT-index is more complex than the h-index in terms of computation, this can be resolved with the aid of a dedicated software program. A description of how the hT-index computation is carried out can be found at the link,^[40] which provides readers with the programming codes to understand how the hT-index is calculated within a second.

4.4. Limitations and suggestions

There are several issues that should be thoroughly examined in further research. The first concern is that the software used to draw the alluvial deposits^[43] is not unique and inevitable. The alluvial (or Sankey) can be drawn online using other software.^[49,53,54] As a result, they are incompatible with the 3 requirements (i.e., weights derived from the arcs in SNA, flows between dimensions extracted backward from the SNA, and nodes aligned and vertically matched to the respective dimension on the x-axis) required for this study.

Second, the dashboards in this study are displayed on Google Maps. Because Google Maps requires a paid project key for use of the cloud platform, these installments are not free of charge. This makes it difficult for other authors to replicate the use in a short period of time.

Third, the calculation of the hT index based on the summation of weights in Ferrers' tableau (i.e., all papers cited in the list) takes a considerable amount of time. Using advanced hardware improvement, the time-consuming process has become trivial, approximately equivalent to computing other bibliometric indices using dedicated software.

Fourth, this IBP was produced online in this study.^[46] However, research achievements are determined by numerous other factors (e.g., the journal impact factor, JIF)^[55] that should also be taken into account when drawing the IBP (e.g., using the JIF-based hT index).

Fifth, in the alluvial plot, only a few dimensions were selected. Other important categories (e.g., research institutes and influential authors in T100ADHD) must also be displayed simultaneously on the alluvial plot. It is recommended that future studies include more dimensions on the x-axis of the alluvial plot.

Finally, although T100ADHDs were primarily extracted from WoS, the results were clearly different in articles from other databases (e.g., Google Scholar, Scopus). T100ADHD should be extracted from more bibliometric databases in future studies.

5. Conclusion

The 2 study objectives were met by demonstrating how to quickly draw the alluvial used for bibliometric analysis and using the alluvial to determine the characteristics of the network among entities of interest in T100ADHD. Bibliometric studies can benefit from the method of drawing the alluvial that is demonstrated in detail with an MP4 video^[42] for readers.

A breakthrough was achieved by reducing Tables and Figures in traditional bibliometric analysis to display network characteristics in articles using the Aluvial plots. Additionally, the use of MeSH terms can be used to classify article subject categories and predict the citations associated with T100ADHD. Bibliometric analyses of 100 top-cited articles can be carried out using the alluvial plot in future studies, and it is not limited to ADHD, as was the case in this study.

Acknowledgment

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

Author contributions

YCT, JW and TWC initiated the research, collected data, conducted the analysis, and wrote the manuscript. CHL contributed to the design of the study and provided critical reviews of the manuscript, and TWC contributed to the interpretation of the results.

Conceptualization: Ya-Ching Tsai, Jian-Wei Wu.

Investigation: Chien-Ho Lin.

Methodology: Tsair-Wei Chien.

References

- Brundage M, Blackford A, Tolbert E, et al. Presenting comparative study PRO results to clinicians and researchers: beyond the eye of the beholder. *Qual Life Res.* 2018;27:75–90.
- Otto E, Culakova E, Meng S, et al. Overview of Sankey flow diagrams: focusing on symptom trajectories in older adults with advanced cancer. *J Geriatr Oncol.* 2022;13:742–6.
- Mica L, Niggli C, Bak P, et al. Development of a visual analytics tool for polytrauma patients: proof of concept for a new assessment tool using a multiple layer Sankey diagram in a single-center database. *World J Surg.* 2020;44:764–72.
- Wang C, Lv T, Deng X. Bibliometric and visualized analysis of China's smart grid research 2008–2018. *Front Res Metr Anal.* 2020;5:551147.
- Lamer A, Laurent G, Pelayo S, et al. Exploring patient path through Sankey diagram: a proof of concept. *Stud Health Technol Inform.* 2020;270:218–22.
- Yu B, Silva CT. VisFlow - web-based visualization framework for tabular data with a subset flow model. *IEEE Trans Vis Comput Graph.* 2017;23:251–60.
- Flannery MA, Culakova E, Canin BE, et al. Understanding treatment tolerability in older adults with cancer. *J Clin Oncol.* 2021;39:2150–63.
- Liu PC, Lu Y, Lin HH, et al. Classification and citation analysis of the 100 top-cited articles on adult spinal deformity since 2011: a bibliometric analysis. *J Chin Med Assoc.* 2022;85:401–8.
- Kuo YC, Chien TW, Kuo SC, et al. Predicting article citations using data of 100 top-cited publications in the journal *Medicine* since 2011: a bibliometric analysis. *Medicine (Baltim).* 2020;99:e22885.
- Ribeca S. Sankey diagrams, parallel sets & alluvial diagrams... What's the Difference? Available at: <https://datavizcatalogue.com/blog/sankey-diagrams-parallel-sets-alluvial-diagrams-whats-the-difference/>. [Access date March 22, 2022].
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Arlington: American Psychiatric Publishing, 2013; pp. 59–65.
- ICD-11-6A05. Attention deficit hyperactivity disorder. Available at: <https://icd.who.int/browse11/l-m/en#/http%3a%2f%2fd.who.int%2fcd%2fentity%2f821852937>. [Access date October 23, 2021].
- Calvo N, Lara B, Serrat L, et al. The role of environmental influences in the complex relationship between borderline personality disorder and attention-deficit/hyperactivity disorder: review of recent findings. *Borderline Personal Disord Emot Dysregul.* 2020;7:1–7.
- Thomas R, Sanders S, Doust J, et al. Prevalence of attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. *Pediatrics.* 2015;135:e994–1001.
- Demontis D, Walters RK, Martin J, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet.* 2019;51:63–75.
- Faraone SV, Asherson P, Banaschewski T, et al. Attention-deficit/hyperactivity disorder. *Nat Rev Dis Primers.* 2015;1:15020.
- Sibley MH, Swanson JM, Arnold LE, et al. Defining ADHD symptom persistence in adulthood: optimizing sensitivity and specificity. *J Child Psychol Psychiatry.* 2017;58:655–62.
- Biederman J, Faraone SV. Attention-deficit hyperactivity disorder. *Lancet.* 2005;366:237–48.
- Skogli EW, Teicher MH, Andersen PN, et al. ADHD in girls and boys—gender differences in coexisting symptoms and executive function measures. *BMC Psychiatry.* 2013;13:298.

- [20] Lin CH, Chien TW, Yan YH. Predicting the number of article citations in the field of attention-deficit/hyperactivity disorder (ADHD) with the 100 top-cited articles since 2014: a bibliometric analysis. *Ann Gen Psychiatry*. 2021;20:6.
- [21] Franke B, Michelini G, Asherson P, et al. Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *Eur Neuropsychopharmacol*. 2018;28:1059–88.
- [22] Matthews M, Nigg JT, Fair DA. Attention deficit hyperactivity disorder. *Curr Top Behav Neurosci*. 2014;16:235–66.
- [23] López-Muñoz F, Alamo C, Quintero-Gutiérrez FJ, et al. A bibliometric study of international scientific productivity in attention-deficit hyperactivity disorder covering the period 1980-2005. *Eur Child Adolesc Psychiatry*. 2008;17:381–91.
- [24] Huang K. Global trends of studies on bone metastasis: a bibliometric and visualization study. *Medicine (Baltim)*. 2022;101:e28761.
- [25] Idriss LT, Hussain M, Khan M, et al. Mapping of global research output in congenital cataracts from 1903 to 2021. *Medicine (Baltim)*. 2021;100:e27756.
- [26] Lei L, Yin S, Meng F, et al. The top 50 most cited articles in carpal tunnel syndrome research: a bibliometrics study. *Medicine (Baltim)*. 2022;101:e28012.
- [27] Li F, Wang S, Yao Y, et al. Visual analysis on the research of mono-carboxylate transporters based on CiteSpace. *Medicine (Baltim)*. 2021;100:e27466.
- [28] Lee YJ, Heo CM, Park S, et al. Top 100 cited articles on hemodialysis: a bibliometric analysis. *Medicine (Baltim)*. 2021;100:e27237.
- [29] Hepting DH. One look is worth a thousand words. Available at: <http://www2.cs.uregina.ca/~hepting/projects/pictures-worth/1921-dec-08.html>. [Access date March 24, 2022].
- [30] Doka G. Sankey diagrams in excel. Available at: <https://www.doka.ch/sankey24.htm>. [Access date March 24, 2022].
- [31] Harris RL. *Information Graphics: A Comprehensive Illustrated Reference*. New York: Oxford University Press, New York. 1999.
- [32] Kan WC, Chou W, Chien TW, et al. The most-cited authors who published papers in JMIR mHealth and uHealth using the authorship-weighted scheme: bibliometric analysis. *JMIR Mhealth Uhealth*. 2020;8:e11567.
- [33] Yang TY, Chen CH, Chien TW, et al. Predicting the number of article citations on the topic of pemphigus vulgaris with the 100 top-cited articles since 2011: a protocol for systematic review and meta-analysis. *Medicine (Baltim)*. 2021;100:e26806.
- [34] Peterson B. Alluvial plots vs Sankey diagrams. Available at: <http://www.datasmith.org/2020/05/02/alluvial-plots-vs-sankey-diagrams/>. [Access date March 20, 2022].
- [35] Rosvall M, Bergstrom CT. Mapping change in large networks. *PLoS One*. 2010;5:e8694.
- [36] Lee YL, Chien TW, Wang JC. Using Sankey diagrams to explore the trend of article citations in the field of bladder cancer: research achievements in China higher than those in the United States. *Medicine*. 2022; in print.
- [37] Chien TW. T100ADHD in this study. Available at <http://www.healthup.org.tw/html100/adhd2022100.htm>. [Access date March 22, 2022].
- [38] Anderson TR, Hankin RKS, Killworth PD. Beyond the Durfee square: enhancing the h-index to score total publication output. *Scientometrics*. 2008;76:577–88.
- [39] Hua PH, Wan JK, Wu JH. A perfect Hirsch-type index? Experiences using the tapered h-Index (hT). *Chinese J Sci Tech Period*. 2010;21:33–7.
- [40] Chien TW. How to compute the hT-index. Available at: <http://www.healthup.org.tw/kpiall/hTindexurl.asp>. [Access date March 22, 2022].
- [41] Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA*. 2005;102:16569–72.
- [42] Chien TW. How to draw the Alluvial diagram in MSExcel. Available at: https://youtu.be/QfD73_QcX58. [Access date March 22, 2022].
- [43] Sankey NATUC. Software for drawing the sankey diagram. Available at: <https://sankeymatic.com/build/>. [Access date January 22, 2022].
- [44] IH. ICite tool. Available at: <https://icite.od.nih.gov/analysis>. [Access date January 7, 2022].
- [45] WoS. Author impact beam plots in web of science author records. Available at: <https://www.youtube.com/watch?v=dcXgx5wxUp4> [access date September 9, 2022].
- [46] Chien TW. How to draw the impact beam plot online. Available at: <http://www.healthup.org.tw/kpiall/Qimpactbeamsplot2.asp>. [Access date January 22, 2022].
- [47] Persson O. BibExcel is designed to assist a user in analysing bibliographic data. Available at: <https://homepage.univie.ac.at/juan.gorraiz/bibexcel/> [access date September 9, 2022].
- [48] Polaczyk GV, Willcutt EG, Salum GA, et al. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *Int J Epidemiol*. 2014;43:434–42.
- [49] Miriamposner. Make an alluvial diagram. Available at: <http://miriamposner.com/classes/dh101f16/tutorials-guides/data-visualization/make-an-alluvial-diagram/>. [Access date March 22, 2022].
- [50] Huang MH, Chi PS. A comparative analysis of the application of h-index, g-index, and a-index in institutional-level research evaluation. *J Library Inform Studi*. 2010;8:1–10.
- [51] Szomszor M. The WoS author impact Beamplots: a new tool for responsible research evaluation. 2022. Available at: <https://clarivate.com/blog/the-web-of-science-author-impact-beamplots-a-new-tool-for-responsible-research-evaluation/>. [Access date January 14, 2022].
- [52] Clarivate. New WOS February 25 release notes. Available at <https://clarivate.com/webofsciencegroup/release-notes/WoS/new-WoS-February-25-release-notes/>. [Access date January 14, 2022].
- [53] Datasketch, Sankey diagram. Available at <https://www.datasketch.co/apps/sankey-diagram/>. [Access date March 23, 2022].
- [54] Datasketch, interactive Sankey diagram. Available at <https://www.datasketch.co/apps/sankey-diagram-interactive/>. [Access date March 23, 2022].
- [55] Yeh JT, Shulruf B, Lee HC, et al. Faculty appointment and promotion in Taiwan's medical schools, a systematic analysis. *BMC Med Educ*. 2022;22:356.