Nexo Revista Científica
Vol. 36, No. 02, pp. 96-109/Marzo 2023

ISSN-E 1995-9516
Universidad Nacional de Ingeniería

http://revistas.uni.edu.ni/index.php/Nexo
https://doi.org/10.5377/nexo.v36i02.16017

Statistical analysis and machine learning in psychoactive substance use: a bibliometric analysis

Análisis estadístico y aprendizaje automático en el consumo de sustancias psicoactivas: un análisis bibliométrico

Kevin Rafael Palomino Pacheco*, Carmen Regina Berdugo Correa

Department of Industrial Engineering. Universidad del Norte. Barranquilla, Colombia.
* krpalomino@uninorte.edu.co

(recibido/received: 01-enero-2023; aceptado/accepted: 27-febrero-2023)

ABSTRACT

Because psychoactive substance use is a topic that has received worldwide attention, this area has added several scientific outcomes. It is essential to conduct a comprehensive analysis comprising as many studies as are available to summarize the separate studies and provide an overall view of how the research field has been evolving over the last few decades. This paper performs a bibliometric analysis using a large dataset of published papers from 2000 to 2021. The study examined 1022 publications from those 20 years. About 79% used statistical analyses, and machine learning techniques were utilized by almost 21%. It is worth mentioning that the publications related to statistical analysis were grouped in the following way: multivariate or univariate statistical analysis (52.4%), Bayesian analysis (21.7%), and spatial analysis (50.5%). There were several key points regarding the overall results of the research. Results illustrated that publications had grown significantly during the last two decades. The majority of the publications come from the United States. In addition, the most prolific authors and journals were identified. Over the last decade, due to advanced technological methods, more research has been focused on enhancing and designing Bayesian techniques for using psychoactive substances.

Keywords: Bibliometric analysis; statistical analysis; machine learning; psychotropic drugs; mental health.

RESUMEN

Dado que el consumo de sustancias psicoactivas es un tema que ha despertado interés a nivel mundial, esta área ha sumado numerosos resultados de carácter científico. Por ello, es esencial llevar a cabo un análisis comprensivo que abarque el mayor número de estudios disponibles para resumir los distintos avances y proporcionar una visión general de cómo ha ido evolucionando el campo de la investigación en las últimas décadas. Este artículo realiza un análisis bibliométrico utilizando un amplio conjunto de datos de artículos publicados entre 2000 y 2021. El estudio examinó 1022 publicaciones de esos 20 años. Alrededor del 79% utilizó técnicas de análisis estadístico, y casi el 21% empleó técnicas de aprendizaje automático. Cabe mencionar que las publicaciones relacionadas con el análisis estadístico se agruparon de la siguiente manera: análisis estadístico multivariante o univariante (52,4%), análisis bayesiano (21,7%) y análisis espacial (50,5%). Hubo varios puntos clave en relación con los resultados generales de la investigación. Los resultados ilustran que las publicaciones han aumentado considerablemente durante las dos últimas décadas. La mayoría de las publicaciones proceden de Estados Unidos. Además, se identificaron los autores y las revistas más prolíficos. En la última década, debido a los avances

tecnológicos, se ha investigado más sobre la potenciación y el diseño de técnicas bayesianas aplicadas al uso de sustancias psicoactivas.

Palabras clave: Análisis bibliométrico; análisis estadístico; aprendizaje automático; psicofármacos; salud mental.

1. INTRODUCTION

Mental health refers mainly to the risks associated with mental problems and disorders, including epilepsy, gender-based violence, domestic abuse, sexual abuse, psychoactive substance use (PSU), schizophrenia, mental retardation, Attention Deficit and Hyperactivity Disorder, along with others. Mental health issues and illnesses are increasingly frequent in the world's population. They are becoming one of the highest disease burdens (Grant et al. 2004; Kessler et al. 1996; Regier et al. 1990). In addition, they constitute a substantial societal cost in terms of losses in productivity, early mortality, rising healthcare spending, criminal justice, well-being costs, and other societal consequences (Goetzel et al. 2003; Sanderson and Andrews 2002; Stewart et al. 2003). The plan entitled 'Comprehensive Mental Health Action 2013-2030' by the World Health Organization (WHO) estimates that the accumulated worldwide impact in terms of economic losses from mental health disorders from 2011 to 2030 will be $16.3 trillion (WHO 2013). Regarding psychoactive substance use (PSU), this term in the literature is equivalent to: 'drug abuse', 'drug use', 'psychotropic substances use', 'psychotropic substances abuse', or 'psychoactive substances abuse'. The WHO has defined Psychoactive Substances Use to mean the consumption of substances that affect mental processes in such a way that if they are not controlled, they can trigger personal, family, social, and educational problems. In brief, PSU is considered a mental disorder caused by the abuse of substances of psychotropic origin. Some commonly abused substances are alcohol, opioids, cannabinoids, sedatives, and hypnotics. Stimulants are also frequently abused (WHO 2018).

The development of research on psychoactive substance use has been oriented towards different approaches. For example, some studies focus on psychoactive substance abuse in teenagers (Ali et al. 2002; Dennis et al. 2004; Grant et al. 2003; Greenmyer et al. 2018; Hingson et al. 2002; Perkins 2002; Schulenberg and Maggs 2002), psychological treatments based on social and emotional learning (SEL) or pharmacological treatments (Durlak et al. 2011; Gudmundsdottir, Weyandt, and Ernudottir 2020; Hostinar, Nusslock, and Miller 2018), and social and geographical implications of drug trafficking (Dolliver, Ericson, and Love 2018; Felson and Staff 2017; Lu, Fang, and Wang 2008; Metternich et al. 2019; Nelson and Obot 2020; Potier et al. 2014). As the PSU has received worldwide interest, this field has produced many scientific results, including previous publications. Therefore, it is crucial to have a comprehensive analysis covering the most significant number of investigations available to consolidate individual studies and show what has been happening in the field over recent decades. Bibliometrics, understood simply as the use of statistical techniques on books and other forms of media, offers a comprehensive information-based analysis of publications with an objective and applied approach (De Bellis 2009; Pritchard 1969). In addition, a bibliometric analysis can give insights into how research is advancing through different procedures, such as performance analysis, which performs a quantitative analysis using citation data (Noyons, Moed, and Luwel 1999; van Raan 2005).

Significant studies conduct a conceptualization, literature review, or bibliometric analysis of psychoactive substance use (Guo et al. 2019; Pallari et al. 2020; Schäfer, Hiemke, and Baumann 2016; Schneider et al. 2014). One of the first articles referring to bibliometric analysis in the use of psychoactive substances is the work of Herrán et al. (1996), who developed a descriptive study based on the review of the journal Atención Primaria on mental health between 1984 and 1995 (Herrán, Artal, and Vázquez-Barquero 1996). Later, some authors studied scientific productivity using bibliometric analysis. For instance, López et al. (2008) conducted a bibliometric analysis of ADHD-related scientific publications from 1980-2005 (Lopez-Munoz et al. 2008); Bramness et al. (2014) compared citation rates within substance abuse

research in Europe with those in the United States from 2001 to 2011 (Bramness et al. 2014); and Sweileh et al. (2014) assessed the productivity of research in the field of substance abuse in Arabic countries employing bibliometric measures (Sweileh et al. 2014). Additionally, researchers are using analytical software packages more often in qualitative studies. In recent work, for instance, Zyoud et all (2017) analyzed publications regarding cocaine intoxication research trends during 1975-2015. They used VOSviewer to show high-frequency terms associated with cocaine toxicity (Zyoud et al. 2017). Thus, despite recent advances in research on psychoactive substance use and the valuable contributions of several authors, there still needs to be more literature regarding statistical and machine learning methods. Also, some of these previous publications need to be updated.

Considering the abundant bibliometric approaches, this paper aims to present a visual overview of the PSU and carry out a comprehensive study of the state of research and advances in this field. In particular, this study presents a bibliometric analysis of scientific publications associated with statistical analysis and machine learning over the last 20 years (2000-2021). The analysis includes publications per year, publications per country, the number of citations per year, most cited publication ranked, most published journal ranked, and most published authors ranked by the total number of publications, among others. A bibliometric analysis of this kind can summarize academic proposals for healthcare policymakers and decision-makers. Likewise, this analysis contributes to the scientific community by establishing a starting point for new collaborations since researchers can learn about the current state of psychoactive substance use research proposals. Regarding its importance for the mental health field, a bibliometric analysis gives an adequate understanding of the different approaches developed and their research directions, identifying different characteristics or aspects of scientific productivity and providing information on the current state of the research agenda as their main gaps.

This manuscript is structured into four main sections, which include this introduction. The second presents methods focusing on how the data have been prepared for analysis. Subsequently, the third part shows the results of this review. Then, the fourth part discusses these results, including opportunities for future work.

2. METHODOLOGY

This paper aimed to conduct a quantitative and visualized analysis of the representative studies related to Statistical Analysis (Bayesian analysis, spatial analysis, or multivariate or univariate statistical analysis) and Machine Learning in PSU through a bibliometric approach. Clarivate Analytics' Web of Science is an electronic indexing service for scientific references that offer comprehensive citation searching. It also provides multiple database access and contains nearly 1.9 billion references cited from more than 171 million records; WoS was therefore selected to obtain the data. The keywords used for the data collection included combinations of “psychoactive substances” and their variant, such as ‘drug abuse’, ‘psychotropic drug’, ‘psychoactive drug’ or ‘psychotropic substances’. Multiple wildcards and searching operators were applied to enhance the precision of the retrieval outputs. Search queries were repeatedly tested and adjusted to find relevant papers. Having found a suitable publication, other search terms were also identified by mapping topic headings and reviewing keywords based on the following queries:

i. *(("Drug Abuse") OR ("Psychoactive Substances") OR ("Psychotropic Drug") OR ("Psychoactive Drug") OR ("Psychotropic Substances") OR (Psychotropic*) OR (Psychoactive*) OR ("Drug Use") OR ("Substance Abuse")) AND (("Machine Learning") OR ("Deep Learning")).*

ii. *(("Drug Abuse") OR ("Psychoactive Substances") OR ("Psychotropic Drug") OR ("Psychoactive Drug") OR ("Psychotropic Substances") OR (Psychotropic*) OR (Psychoactive*) OR ("Drug Use") OR ("Substance Abuse")) AND (("Spatial Analysis") OR ("Spatial Statistic") OR ("Spatial Statistical") OR (Geostatistical*) OR (Geostatistic*) OR (Geoanalysis*)).*

iii. *(("Drug Abuse") OR ("Psychoactive Substances") OR ("Psychotropic Drug") OR ("Psychoactive Drug") OR ("Psychotropic Substances") OR (Psychotropic*) OR (Psychoactive*) OR ("Drug*

Use") OR ("Substance Abuse")) AND (("Bayesian Analysis") OR ("Bayesian Statistic") OR ("Bayesian Statistical") OR (Bayesian) NOT (("Spatial Analysis") OR ("Spatial Statistic") OR ("Spatial Statistical") OR (Geostatistical*) OR (Geostatistic*) OR (Geoanalysis*))).*

iv. *(("Drug Abuse") OR ("Psychoactive Substances") OR ("Psychotropic Drug") OR ("Psychoactive Drug") OR ("Psychotropic Substances") OR (Psychotropic*) OR (Psychoactive*) OR ("Drug Use") OR ("Substance Abuse")) AND (("Statistical Analysis") OR ("Statistic Analysis") NOT (("Bayesian Analysis") OR ("Bayesian Statistic") OR ("Bayesian Statistical") OR (Bayesian*) OR ("Spatial Analysis") OR ("Spatial Statistic") OR ("Spatial Statistical") OR (geostatistical*) OR (geostatistic*) OR (Geoanalysis*))).*

Our search was strictly limited to peer-reviewed articles from journals published in English from 2000 to November 2021. References were excluded based on title if they were not in the relevant subject area: Published articles in which statistical analysis (Bayesian analysis, spatial analysis, or multivariate or univariate statistical analysis) or machine learning was used in Psychoactive Substances. For each publication, both authors completed the data collection. The discussion continued if the authors disagreed until a consensus was reached. To encompass as much information as possible, we exported complete records for each retrieval item in ".csv" file format in WoS. RStudio software was used to verify and delete duplicate and non-related items by hand. Different data types were extracted, including total publications per year, total publications per country, number of citations per year for each publication, the most cited publication ranked, most published journal ranked, and most published authors ranked, among others.

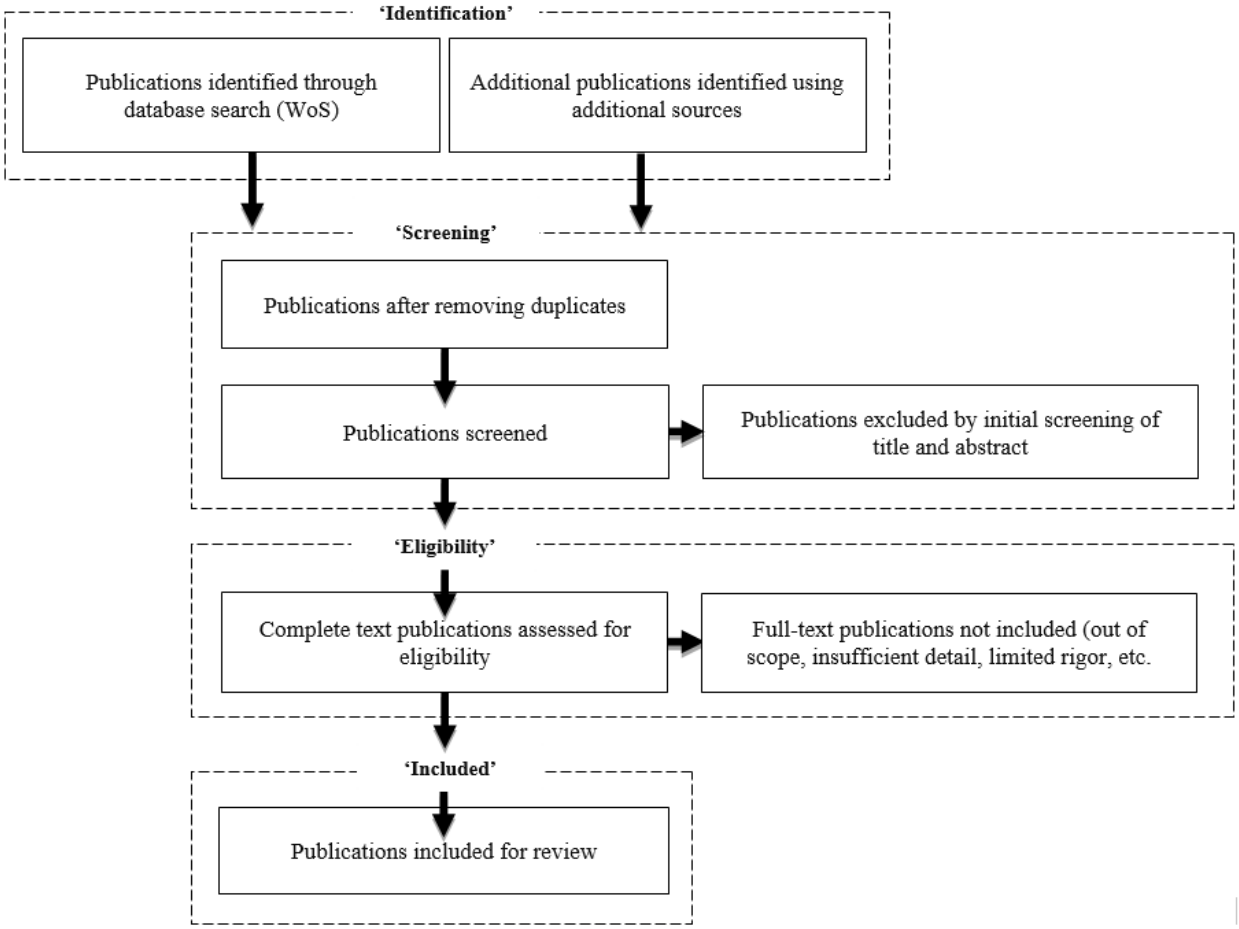

Figure 1. Bibliometric analysis flow chart.

A detailed description of the publication selection process can be found in the flow chart in Figure 1. Using the filtering and search scheme, 1,022 publications were identified for review. Technically, numerous tools and computer software packages are available to perform the bibliometric process. Among these tools, RStudio is an open-source software for data science developed by Joseph J. Allaire. RStudio was robust enough to conduct approximately all the bibliometric analyses. In addition, VOSviewer executed visualized and interactive features for easy understanding of patterns, including composing automated labeling of clusters based on cited publication terms. Further criteria adopted were: Where available, additional references were searched through a second search of the reference lists in key publications. Experts in this field were also asked to confirm that all key articles had been included. This aspect was especially critical considering the search terms involved in this subject field: these were commonly used words unrelated to the subject matter, which resulted in a high volume of irrelevant articles. All articles of the relevant subject area were then reviewed using a data extraction sheet, and key findings were reviewed. Finally, papers and book chapters that were distinctly publication reviews were not included; those that were relevant are discussed in the introductory section.

3. RESULTS

Table 1 indicates the number of academic publications in English between 2000 and November 2021 using four analytical techniques. 1,022 publications were examined in those twenty years (all references are provided as supplementary material). About 79% (n = 808) used statistical analyses. Machine learning techniques were utilized by almost 21% (n = 214). It is worth mentioning that the publications related to Statistical Analysis were grouped as follows: Multivariate or Univariate Statistical Analysis - MUSA (n = 536, 52.4%), Bayesian Analysis (n = 222, 21.7%), and Spatial Analysis (n = 50,5%).

Table 1. Total number of publications.

Field	Number of publications (N = 1,022)
1. Statistical Analysis	808 (79.1%)
1.1. Bayesian Analysis	222
1.2. Spatial Analysis	50
1.3. Multivariate or Univariate Statistical Analysis	536
2. Machine Learning	214 (20.9%)

In this way, in the last twenty years, publications related to Bayesian analysis on psychoactive substances have grown steadily, peaking in 2019 with 25 papers. The annual number of publications on psychoactive substances using Spatial Analysis averages 7.17 publications per year, with a peak of 7 in 2019. The number of publications using multivariate or univariate statistical analysis (MUSA) has also increased. The annual number of publications on psychoactive substances using MUSA was 19.34, with a slight increase from 2008 onwards and a peak of 63 publications in 2020. Publications using Machine Learning show an uneven skew to the distribution. From 2001 to 2014, the number of publications was relatively flat; from 2016-2021, the number had a higher output. Thus, publications present an average annual growth rate of 33.3% and have risen considerably since 2014, with a peak of 54 in 2019. In general, although the use of statistical analysis increased over the years, machine learning was the technique that showed the highest annual growth rate. The increase in the total of publications reflects an expansion of global scientific inquiry into this field. Figure 1 presents the number of publications per year.

Bibliometrix R-package was used to calculate and plot the performance of authors (in terms of the total publications) in descending order. Figure 3 below outlines each approach's top ten contributing authors and the number of publications.

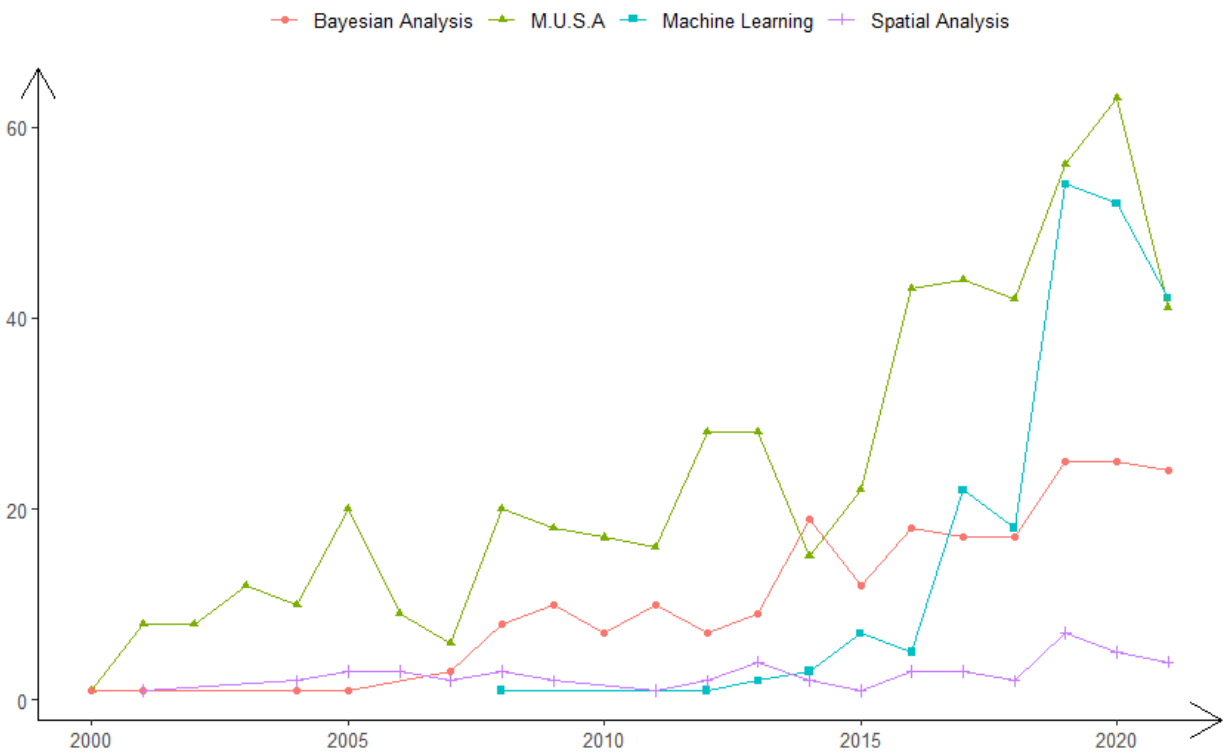

Figure 2. Publications per year.

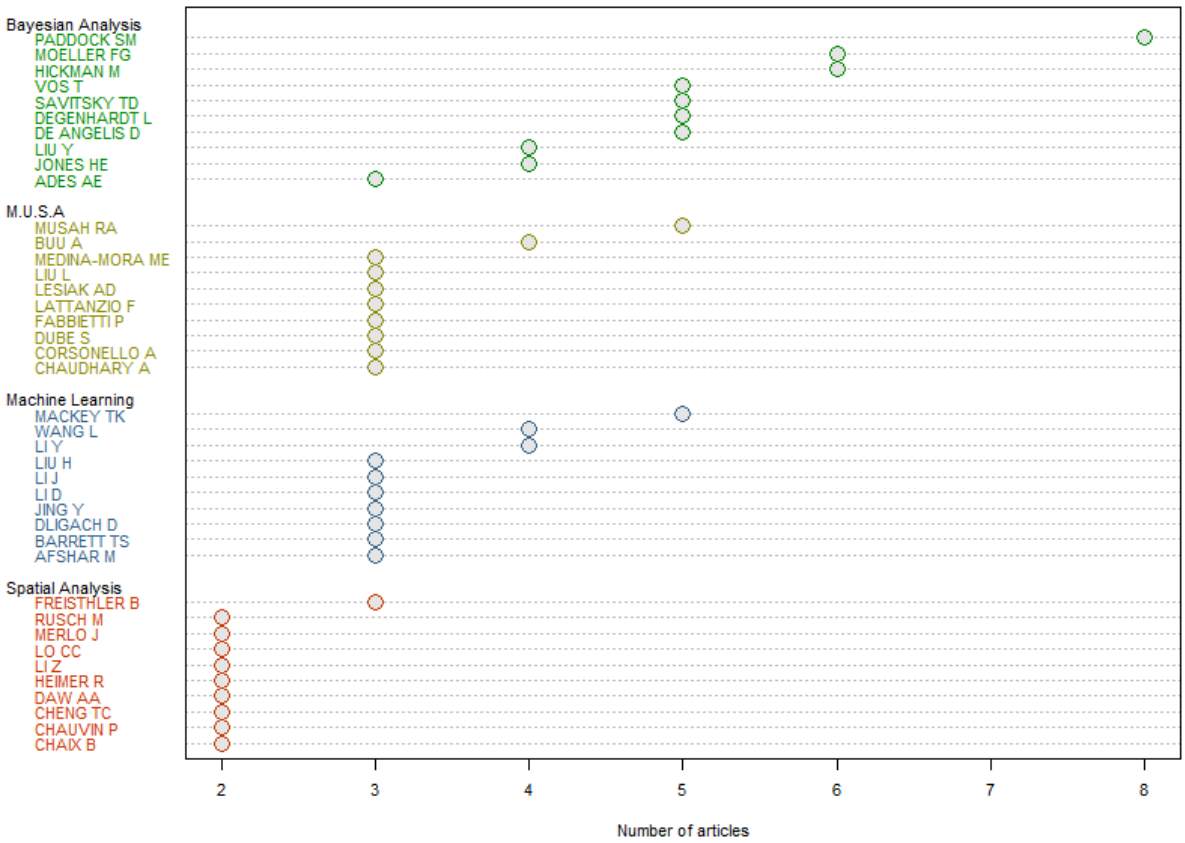

Figure 3. Performance of authors.

The authors with the highest publication numbers were Paddock, Moeller, and Hickman (8, 6, and 6 articles published, respectively) using Bayesian techniques on psychoactive substances research; their research background is related to substance abuse treatment in therapy groups and neuroimaging and genetic studies concerning drug dependence. As for Spatial analysis, Freisthler was the author with more publications (3 scientific articles in total); their research is related to drug market operations and child abuse geography; for this approach, there were fewer than three publications by other authors. It appears that Musah and Buu dominated the list of contributing authors with the most significant number of publications using multivariate or univariate statistical analysis in psychoactive substance use with 5 and 4 publications, respectively; their research interests are in mass spectrometry for the detection and identification of psychoactive plants, and genome-wide association studies for drug dependence. Regarding the authors leading the list in machine learning, Mackey and Wang have published the most, having a total of 5 and 4 publications, respectively; their research interests are associated with predicting the severity of substance use severity from childhood through adulthood. Comparing the results, it should be noted that the size of the active window necessary to be considered the primary active author is different in each case.

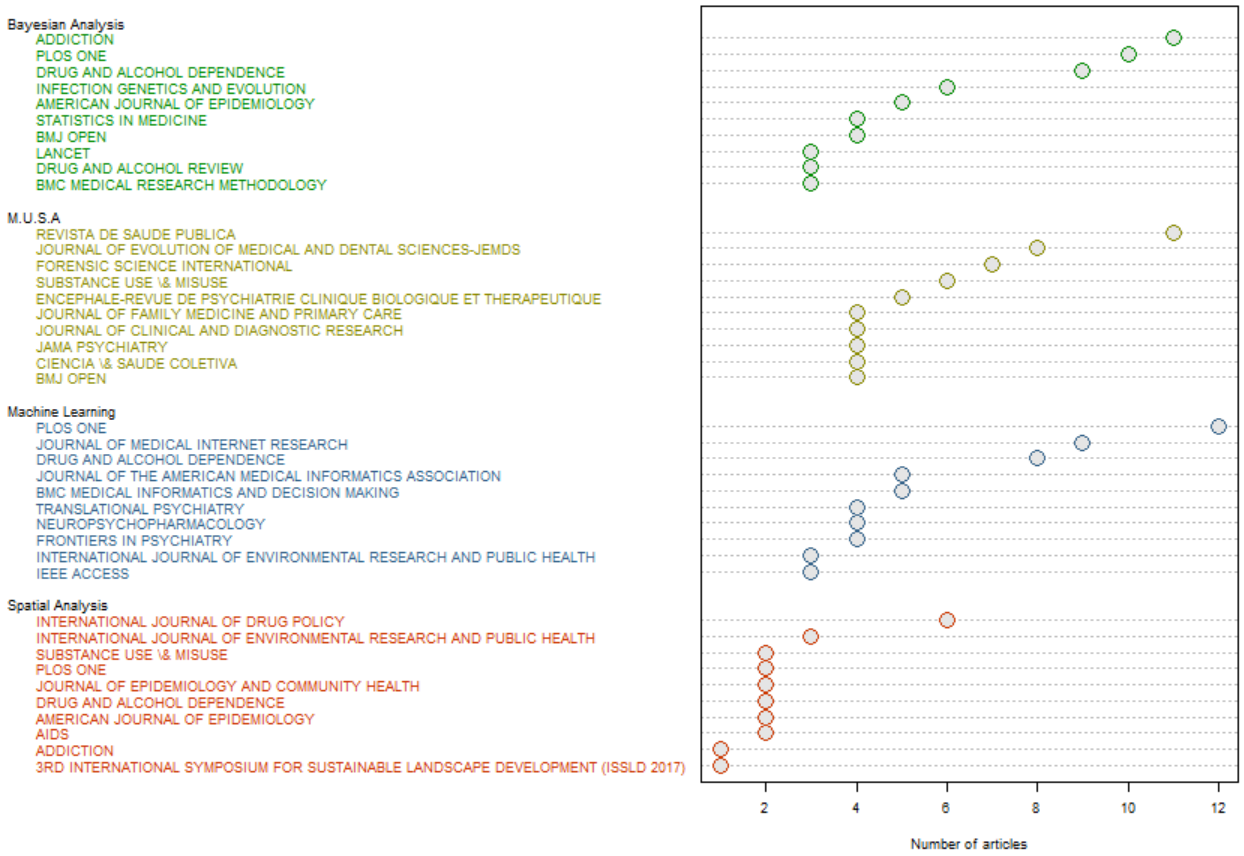

Figure 4. Journal performance.

Analyzing journal performance results can help determine which journals have contributed the most scientific papers, generated significant interest, and which journal a manuscript should be submitted to. Figure 4 shows the 10 most fruitful journals in decreasing order of the number of publications. 'Plos One' ranks first with 11 publications, followed by 'Addiction' with 10 publications using Bayesian analysis approaches. Regarding spatial analysis, the most productive journals are the 'International Journal of Drug Policy' with 6 articles, and the 'International Journal of Environmental Research and Public Health' with only 3 in total. As to the multivariate or univariate statistical analysis, the journal 'Revista de Saúde Pública' ranked first with 11 registered publications, followed by the 'Journal of Evolution of Medical and

Dental Sciences' with 8 published papers. In Machine Learning, once again ‘Plos One’ is the leader of the ranking with 12 publications, with the ‘Journal of Medical Internet Research’ following with 9 publications. From a broad point of view, ‘Plos One’ is the most productive journal, reaching a total of 24 publications in all the approaches included in this review, followed by the ‘Drug and Alcohol Dependence’, and the journal ‘Addiction’, with 19 and 12 publications, respectively. It is relevant to highlight that the journal ‘Plos One’ has the highest H-index (300), followed by Revista de Saúde Pública’ with 182.

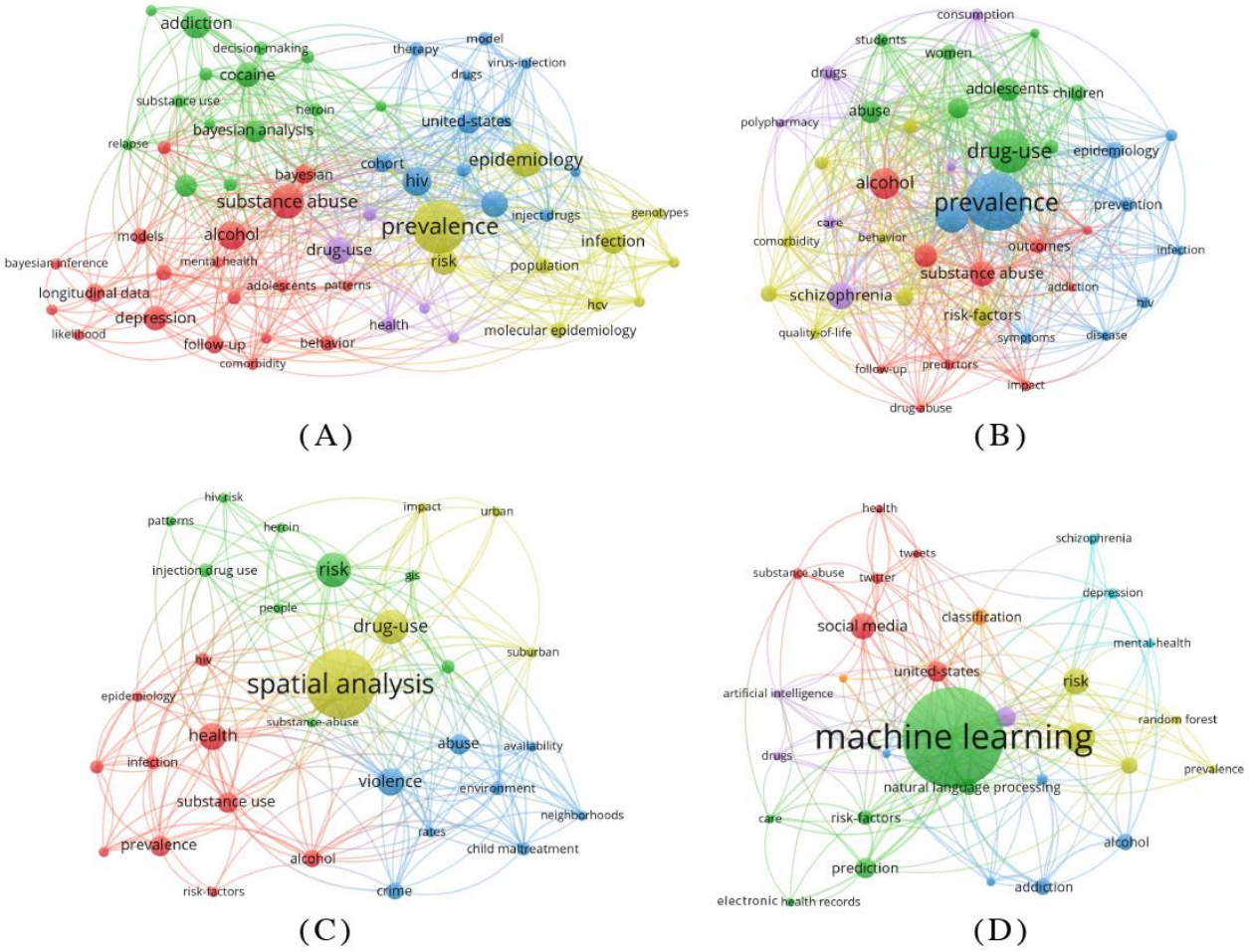

Figure 5. Network visualization map of author keywords.

Figure 5 is the network analysis graph focusing on the keyword combination. The map circles indicate the keywords used in the publications extracted. The wider the circle, the higher the level of keyword frequency. Fig.5(A) shows some clusters of keyword occurrences associated with Bayesian analysis. Overall, Bayesian techniques applied in studies of psychoactive substance use aim to find epidemiological patterns through models to analyze prevalence, comorbidity, and mortality related to drugs in the population; Researchers seem to use longitudinal data to try to understand within-sample changes over time and how those changes lead to outcomes detected in later waves of the data. Furthermore, in Fig. 5(B), the network map for trends based on keyword analysis in multivariate or univariate statistical analysis is presented. As can be seen, these techniques are most associated with drug use and abuse, addiction, and prevalence in adolescents. Alcohol consumption, the risk of developing schizophrenia, and the use of predictive models are highlighted. Similarly, Fig. 5(C) presents some clusters of keywords associated with spatial analysis. The research emphasizes the risks and patterns of PSU and the presence of crime, violence, and child abuse in urban or suburban environments. Also critical is the use of

Geographic Information Systems and the analysis of epidemiological prevalence factors. As shown in Fig. 5(D), techniques such as Social Media Monitoring, Natural Language Processing, Random Forest, and Deep learning are highlighted in the field of machine learning. The prediction and prevalence associated with risk factors and mental health problems, including depression and schizophrenia, are also essential. The United States seems to be where these studies are most widely applied.

Another important aspect was the mapping of the most productive countries. Figure 6 presents the total number of publications and the most contributing countries in the past 20 years. Spatial calculations were executed using the RStudio packages "tmap", "rgeos", and "rgdal". Fig. 6(A) shows the most productive countries publishing the most significant articles associated with Bayesian analysis on psychoactive substance use from 2000 to 2021. As can be seen, the United States generates the greatest number of publications (73; 42%), followed by the United Kingdom and Australia, with 17 and 14 articles published, respectively. In the same way, the most productive country in the publishing of articles associated with multivariate or univariate analysis of psychoactive substance use is the United States, with 95 publications, which represents 22% of the total number of articles published, followed by Brazil (39; 8%) and India (24; 5%), as shown in Fig. 6(B). In Fig. 6(C), the United States also leads the ranking of countries with the most publications on psychoactive substance research using spatial analysis, with 22 publications representing 54% of the total number of articles published in this field. The list is preceded by Canada (4; 9%) and China (3; 7%). Last but not least, as seen from Fig. 6(D), the United States is again the country that contributes with the most significant number of scientific publications (86;66%), followed by China (73; 42%) and India (5;4%) on psychoactive substances research using machine learning.

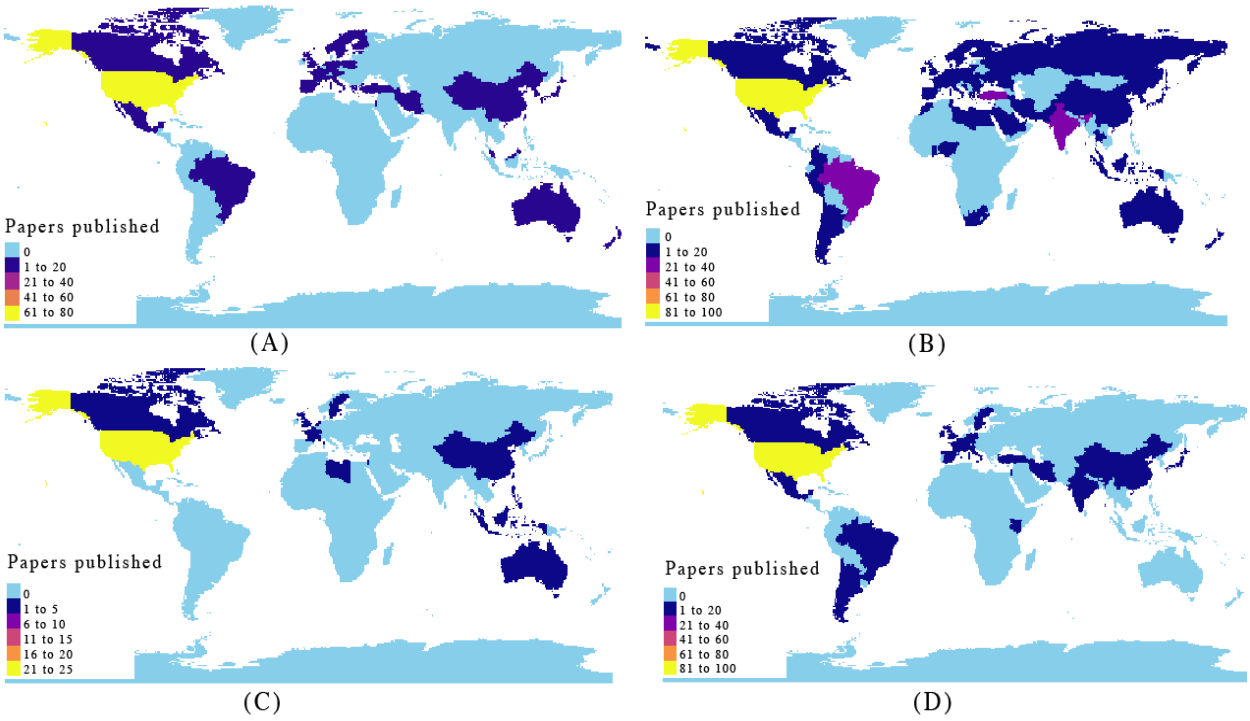

Figure 6. Performance of countries.

In addition, the 10 references with the highest bursts of citations from 2000 to 2021 were also exported, as shown in Table 2. The academic performance of a particular area of research can be described by the publications that received the most outstanding increment of references, i.e., the bursts of citations. Citation burst suggests the probability that the academic community has paid or is paying particular attention to the underlying contribution. In the case of Machine learning, Wager et al. (2013) led the

ranking with their work entitled "An fMRI-Based Neurologic Signature of Physical Pain", which registered 534 citations with 66.7 citations on average per year. It is important to mention that the most cited techniques in this group of top publications are ‘support vector machine’, ‘random forest’, ‘natural language processing’, and ‘social media monitoring’. Regarding Bayesian analysis, the most prominent publication is "Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study" by Whiteford et al (2010). This work records a total of 2337 citations, averaging 292 per year. From a general point of view, Bayesian meta-regression and Bayesian hierarchical models are the most attractive approaches.

For Spatial Analysis, the work of Shannon et al. (2008) entitled "Mapping violence and policing as an environmental-structural barrier to health service and syringe availability among substance-using women in street-level sex work" is the most cited (162) with a mean of 12.5 citations per year. It should be noted that spatial distributions, spatial regression models, spatial scan statistics, variograms, and social mapping are the most widely used techniques to address the problem of psychoactive substance use. Finally, as far as Multivariate or Univariate Statistical Analysis is concerned, the most striking article is "Structural abnormalities in the brains of human subjects who use methamphetamine", which was carried out by Thompson et al. (2004). It is worth mentioning that techniques such as independent group T-Test, descriptive analysis correlations, logistic regression, and confirmatory factor analysis are associated with the most cited papers. In this way, academic interest in these highly cited articles led to forming the main representative approaches. Knowledge of these papers with solid bursts of citations can provide a quick insight into developing new research interests and the continuing evolution of treatment for psychoactive substance use.

Table 1. Top 10 most cited publications.

		Ranking	Approach	Times cited	Times cited per year
1	**Machine learning**	(Wager et al. 2013)	LASSO-PCR (Least Absolute Shrinkage and Selection Operator-Regularized Principal Components Regression)	534	66.75
2		(Lee 2013)	Model-Based Reinforcement Learning	63	7.88
3		(Henriksson et al. 2015)	Natural Language Processing	41	6.83
4		(Conway et al. 2016)	Social Media, Big Data and Natural Language Processing	38	7.60
5		(Squeglia et al. 2017)	Random Forest	37	9.25
6		(Arnold et al. 2012)	Ensemble Machine Learning with Random Forests	35	3.89
7		(Bedi et al. 2014)	Natural Language Toolkit and Support Vector Machines	31	4.43
8		(Katsuki et al. 2015)	Support Vector Machines	27	4.50
9		(Struyf et al. 2008)	Support Vector, Nearest Shrunken Centroids, Decision Trees, Nave Bayes, and Nearest Neighbor	25	1.92
10		(Alvaro et al. 2015)	NPL, SVM, GLM, C50, Naive Bayes, Bayesian Generalized Linear Model, and Multi-Layer Perceptron	24	4.00
1	**Bayesian Analysis**	(Whiteford et al. 2013)	Disease Modeling – Meta-Regression, Bayesian Meta-Regression	2337	292.12
2		(Woolcott et al. 2009)	Updated Bayesian Odds Ratios	591	49.25
3		(Degenhardt et al. 2013)	Bayesian Meta-Regression Technique (DisMod-MR)	392	49.00
4		(Collins et al. 2015)	Bayesian Model Averaging	140	23.33
5		(Sung et al. 2004)	Hierarchical Bayesian Framework, Gibbs Sampling Approach	118	6.94
6		(Schwan et al. 2010)	Bayesian Data-Mining Algorithm	88	8.00
7		(Rivkees et al. 2010)	Gamma-Poisson Shrinker (MGPS) Data Mining Algorithm, Bayes Geometric Mean	84	7.64
8		(Degenhardt et al. 2014)	Bayesian Meta-Regression Technique (DisMod-MR)	74	10.57
9		(Bowman et al. 2008)	Bayesian Extension of Voxel-Level Analyses, Markov Chain Monte Carlo, Bayesian Hierarchical Model	70	5.38
10		(O’Brien et al. 2014)	Bayesian Inference Models	64	9.14
1	**Spatial Analysis**	(Shannon et al. 2008)	Social Mapping	162	12.46
2		(Chaix et al. 2005)	Spatial Analytical Perspective, Geo-additive Models, Multilevel Approach	123	7.69
3		(Freisthler 2004)	Spatial Regression Models	123	7.24
4		(Prosser et al. 2006)	Visual-Spatial Analysis	99	6.60

5		(Banta-Green et al. 2009)	Spatial Epidemiology, Spatial Distribution	88	7.33
6		(Freisthler et al. 2005)	Spatial Regression Techniques	87	5.44
7		(Dovey et al. 2001)	Socio-Spatial Analysis	84	4.20
8		(Chaix et al. 2006)	Spatial Distributions, Spatial Scan Statistic	52	3.47
9		(Heimer et al. 2008)	Spatial Distribution	36	2.77
10		(Bass et al. 2004)	Variograms (Spatial Dependence)	24	1.41
1	**MUSA**	(Thompson et al. 2004)	Student's T Test, Descriptive Analysis, Confidence Intervals	440	25.88
2		(Mowbray et al. 2003)	Internal Consistency Reliability, CFA, Cluster Analysis, Pearson Correlations, Coefficient Kappa,	414	23.00
3		(Fazel et al. 2008)	Confidence Interval, Cochran's Q Test, I^2 Statistic	308	23.69
4		(Shoji et al. 2004)	Logistic Regression Analysis, Descriptive Analysis	269	15.82
5		(Maas et al. 2006)	Proportional Odds Model	165	11.00
6		(Peet et al. 2005)	Correlation	163	10.19
7		(Gilbert et al. 2006)	Descriptive Analysis	149	9.93
8		(Neubert et al. 2004)	Fisher's Exact Test, Chi-Quadrat Test, Descriptive Analysis	106	6.24
9		(Wazaify et al. 2005)	Descriptive Analysis	98	6.12
10		(Merson et al. 2000)	Independent Group T-Test and The Paired T-Test.	92	4.38

4. DISCUSSION AND CONCLUSION

Based on the data set, including 1,022 publications from Web of Science from 2000 to 2021, this paper conducted a complete bibliometric analysis of statistical analysis and machine learning in psychoactive substance use. Several important facts about research performance were observed. First, results illustrated that annual articles grew considerably in the past two decades. The United States notably contributed to the most significant number of publications, followed by China, Brazil, and India. Furthermore, the most prolific authors and journals were identified. In the last decade, due to advanced technological methods, most research has steered toward attempts to enrich and design Bayesian techniques on psychoactive substance use. Few spatial analysis papers appear in journals, likely due to researchers' perceptions concerning spatial data access. Additionally, it can be concluded that machine learning and multivariate or univariate statistical analysis will remain a critical part of understanding psychoactive substances. Thus, the psychoactive substance use study analysis is meaningful to researchers; for example, deciding which researcher to follow or co-author with, determining which journal to focus on or submit manuscripts to, and selecting which country to enrich collaboration or exchanges with. Also, the keyword network analysis showed can help researchers understand the landscape of psychoactive substance use and set up future research directions. In brief, the bibliometric analysis provides solid evidence that PSU is becoming gradually accepted, validated, and embraced in broader geographical regions, research fields, and periods.

Different techniques and approaches used across all the publications are further complications. Publications involved in Bayesian analysis mostly belonged to the epidemiological models. In contrast, those involved in multivariate or univariate statistical analysis belonged to predictive models and techniques associated with drug use and abuse, addiction, and prevalence, mainly in adolescents. Regarding spatial analysis research associated with the psychoactive substance use, risks and patterns of PSU, and the presence of crime, violence, and child abuse in urban or suburban environments were the most frequently studied; while social media monitoring, natural language processing, random forest, and deep learning applications were most frequently involved in machine learning. We did not find the quality of publications to be related to the manuscript length and the number of studies included. Although this article followed the bibliometric analysis process, there are certain limitations. The principal limitation is that this research was restricted to papers published in WoS; still, it is possible to argue that results might also be helpful to other journals that are not explicitly connected with this database. Methodologically, in addition to bibliometric analysis, numerous approaches are available to review publications, such as systematic review and meta-analysis. Those methods might present different outcomes. Comparisons between methods might produce results complementary to the present findings and give more clear

guidance in the psychoactive substance use domain. Finally, Further investigations are needed to clarify specific features of the design and approach of these publications, outcomes, and follow-up research. Our review also shows that publications in this field still need to be sparser to draw more specific conclusions. Different samples used in different studies, heterogeneity of treatment duration, multiple exclusion factors, and the high number of studies reporting control groups generalize results from one study to another extremely complex. Future efforts should address questions that focus on homogeneous groups and address the results in depth.

REFERENCIAS

Ali, R., E. Awwad, T. F. Babor, F. Bradley, T. Butau, M. Farrell, MLOS Formigoni, R. Isralowitz, R. B. de Lacerda, J. Marsden, B. McRee, M. Monteiro, H. Pal, M. Rubio-Stipec, J. Vendetti, and W. H. O. ASSIST Working Grp. 2002. "The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): Development, Reliability and Feasibility." *ADDICTION* 97(9):1183–94.

De Bellis, N. 2009. *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Scarecrow Press.

Bramness, Jorgen G., Beate Henriksen, Olle Person, and Karl Mann. 2014. "A Bibliometric Analysis of European versus USA Research in the Field of Addiction. Research on Alcohol, Narcotics, Prescription Drug Abuse, Tobacco and Steroids 2001-2011." *EUROPEAN ADDICTION RESEARCH* 20(1):16–22. doi: 10.1159/000348260.

Dennis, M., S. H. Godley, G. Diamond, F. M. Tims, T. Babor, J. Donaldson, H. Liddle, J. C. Titus, Y. Kaminer, C. Webb, N. Hamilton, and R. Funk. 2004. "The Cannabis Youth Treatment (CYT) Study: Main Findings from Two Randomized Trials." *JOURNAL OF SUBSTANCE ABUSE TREATMENT* 27(3):197–213. doi: 10.1016/j.jsat.2003.09.005.

Dolliver, Diana S., Steven P. Ericson, and Katherine L. Love. 2018. "A Geographic Analysis of Drug Trafficking Patterns on the TOR Network." *GEOGRAPHICAL REVIEW* 108(1):45–68. doi: 10.1111/gere.12241.

Durlak, Joseph A., Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger. 2011. "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions." *CHILD DEVELOPMENT* 82(1):405–32. doi: 10.1111/j.1467-8624.2010.01564.x.

Felson, Richard B., and Jeremy Staff. 2017. "Committing Economic Crime for Drug Money." *CRIME & DELINQUENCY* 63(4):375–90. doi: 10.1177/0011128715591696.

Goetzel, R. Z., K. Hawkins, R. J. Ozminkowski, and S. H. Wang. 2003. "The Health and Productivity Cost Burden of the 'top 10' Physical and Mental Health Conditions Affecting Six Large US Employers in 1999." *JOURNAL OF OCCUPATIONAL AND ENVIRONMENTAL MEDICINE* 45(1):5–14. doi: 10.1097/00043764-200301000-00007.

Grant, B. F., D. A. Dawson, F. S. Stinson, P. S. Chou, W. Kay, and R. Pickering. 2003. "The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): Reliability of Alcohol Consumption, Tobacco Use, Family History of Depression and Psychiatric Diagnostic Modules in a General Population Sample." *DRUG AND ALCOHOL DEPENDENCE* 71(1):7–16. doi: 10.1016/S0376-8716(03)00070-X.

Grant, B. F., F. S. Stinson, D. A. Dawson, S. P. Chou, M. C. Dufour, W. Compton, R. P. Pickering, and K. Kaplan. 2004. "Prevalence and Co-Occurrence of Substance Use Disorders and Independent Mood and Anxiety Disorders - Results from the National Epidemiologic Survey on Alcohol and Related Conditions." *ARCHIVES OF GENERAL PSYCHIATRY* 61(8):807–16. doi: 10.1001/archpsyc.61.8.807.

Greenmyer, Jacob R., Marilyn G. Klug, Cassondra Kambeitz, Svetlana Popova, and Larry Burd. 2018. “A Multicountry Updated Assessment of the Economic Impact of Fetal Alcohol Spectrum Disorder: Costs for Children and Adults.” *JOURNAL OF ADDICTION MEDICINE* 12(6):466–73. doi: 10.1097/ADM.0000000000000438.

Gudmundsdottir, Bergljot Gyda, Lisa Weyandt, and Gabriela Bryndis Ernudottir. 2020. “Prescription Stimulant Misuse and ADHD Symptomatology Among College Students in Iceland.” *JOURNAL OF ATTENTION DISORDERS* 24(3, SI):384–401. doi: 10.1177/1087054716684379.

Guo, F., W. Lv, L. Liu, T. Wang, and V. G. Duffy. 2019. “Bibliometric Analysis of Simulated Driving Research from 1997 to 2016.” *Traffic Injury Prevention* 20(1):64–71. doi: 10.1080/15389588.2018.1511896.

Herrán, A., J. Artal, and J. L. Vázquez-Barquero. 1996. “Mental Health in Primary Care: A Bibliometric Study.” *Atencion Primaria / Sociedad Española de Medicina de Familia y Comunitaria* 18(9):502–6.

Hingson, R. W., T. Heeren, R. C. Zakocs, A. Kopstein, and H. Wechsler. 2002. “Magnitude of Alcohol-Related Mortality and Morbidity among US College Students Ages 18-24.” *JOURNAL OF STUDIES ON ALCOHOL* 63(2):136–44. doi: 10.15288/jsa.2002.63.136.

Hostinar, Camelia E., Robin Nusslock, and Gregory E. Miller. 2018. “Future Directions in the Study of Early-Life Stress and Physical and Emotional Health: Implications of the Neuroimmune Network Hypothesis.” *JOURNAL OF CLINICAL CHILD AND ADOLESCENT PSYCHOLOGY* 47(1):142–56. doi: 10.1080/15374416.2016.1266647.

Kessler, R. C., C. B. Nelson, K. A. McGonagle, M. J. Edlund, R. G. Frank, and P. J. Leaf. 1996. “The Epidemiology of Co-Occurring Addictive and Mental Disorders: Implications for Prevention and Service Utilization.” *American Journal of Orthopsychiatry* 66(1):17–31. doi: 10.1037/h0080151.

Lopez-Munoz, Francisco;, Cecilio; Alamo, Francisco Javier; Quintero-Gutierrez, and Pilar; Garcia-Garcia. 2008. “A Bibliometric Study of International Scientific Productivity in Attention-Deficit Hyperactivity Disorder Covering the Period 1980-2005.” *EUROPEAN CHILD & ADOLESCENT PSYCHIATRY* 17(6):381–91. doi: 10.1007/s00787-008-0680-1.

Lu, Lin, Yuxia Fang, and Xi Wang. 2008. “Drug Abuse in China: Past, Present and Future.” *CELLULAR AND MOLECULAR NEUROBIOLOGY* 28(4):479–90. doi: 10.1007/s10571-007-9225-2.

Metternich, Sonja, Siegfried Zoerntlein, Torsten Schoenberger, and Carolin Huhn. 2019. “Ion Mobility Spectrometry as a Fast Screening Tool for Synthetic Cannabinoids to Uncover Drug Trafficking in Jail via Herbal Mixtures, Paper, Food, and Cosmetics.” *DRUG TESTING AND ANALYSIS* 11(6):833–46. doi: 10.1002/dta.2565.

Nelson, Ediomo-Ubong, and Isidore Obot. 2020. “Beyond Prohibition: Responses to Illicit Drugs in West Africa in an Evolving Policy Context.” *DRUGS AND ALCOHOL TODAY*. doi: 10.1108/DAT-07-2019-0033.

Noyons, E. C. M., H. F. Moed, and M. Luwel. 1999. “Combining Mapping and Citation Analysis for Evaluative Bibliometric Purposes: A Bibliometric Study.” *Journal of the American Society for Information Science* 50(2):115–31. doi: https://doi.org/10.1002/(SICI)1097-4571(1999)50:2<115::AID-ASI3>3.0.CO;2-J.

Pallari, E., T. Soukup, A. Kyriacou, and G. Lewison. 2020. “Assessing the European Impact of Alcohol Misuse and Illicit Drug Dependence Research: Clinical Practice Guidelines and Evidence-Base Policy.” *Evidence-Based Mental Health* 23(2):67–76. doi: 10.1136/ebmental-2019-300124.

Perkins, H. W. 2002. “Social Norms and the Prevention of Alcohol Misuse in Collegiate Contexts.” *JOURNAL OF STUDIES ON ALCOHOL* (14):164–72. doi: 10.15288/jsas.2002.s14.164.

Potier, Chloe, Vincent Laprevote, Francoise Dubois-Arber, Olivier Cottencin, and Benjamin Rolland. 2014. “Supervised Injection Services: What Has Been Demonstrated? A Systematic Literature Review.” *DRUG AND ALCOHOL DEPENDENCE* 145:48–68. doi: 10.1016/j.drugalcdep.2014.10.012.

Pritchard, Alan. 1969. “Statistical Bibliography or Bibliometrics?” *Journal of Documentation* 25:348–49.
van Raan, Anthony F. J. 2005. “Measuring Science BT - Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems.” Pp. 19–50 in, edited by H. F. Moed, W. Glänzel, and U. Schmoch. Dordrecht: Springer Netherlands.

Regier, Darrel A., Mary E. Farmer, Donald S. Rae, Ben Z. Locke, Samuel J. Keith, Lewis L. Judd, and Frederick K. Goodwin. 1990. “Comorbidity of Mental Disorders With Alcohol and Other Drug Abuse: Results From the Epidemiologic Catchment Area (ECA) Study.” *JAMA* 264(19):2511–18. doi: 10.1001/jama.1990.03450190043026.

Sanderson, K., and G. Andrews. 2002. “Prevalence and Severity of Mental Health-Related Disability and Relationship to Diagnosis.” *PSYCHIATRIC SERVICES* 53(1):80–86. doi: 10.1176/appi.ps.53.1.80.

Schäfer, A., C. Hiemke, and P. Baumann. 2016. “Consensus Guideline for Therapeutic Drug Monitoring in Psychiatry (2004): Bibliometric Analysis of Citations for the Period 2004-2011.” *Nordic Journal of Psychiatry* 70(3):202–7. doi: 10.3109/08039488.2015.1080296.

Schneider, D. R., A. Vidal-Infer, M. Bolaños-Pizarro, R. Aleixandre-Benavent, F. J. B. Cañigral, and J. C. Valderrama-Zurián. 2014. “Scientific Collaboration between Latin America and the European Union (2001-2010) on Drug Abuse from the ISI Web of Science.” *Salud Mental* 37(3):199–210.

Schulenberg, J. E., and J. L. Maggs. 2002. “A Developmental Perspective on Alcohol Use and Heavy Drinking during Adolescence and the Transition to Young Adulthood.” *JOURNAL OF STUDIES ON ALCOHOL* (14):54–70. doi: 10.15288/jsas.2002.s14.54.

Stewart, W. F., J. A. Ricci, E. Chee, S. R. Hahn, and D. Morganstein. 2003. “Cost of Lost Productive Work Time among US Workers with Depression.” *JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION* 289(23):3135–44. doi: 10.1001/jama.289.23.3135.

Sweileh, W. M., S. H. Zyoud, S. W. Al-Jabi, and A. F. Sawalha. 2014. “Substance Use Disorders in Arab Countries: Research Activity and Bibliometric Analysis.” *Substance Abuse Treatment, Prevention, and Policy* 9:33. doi: 10.1186/1747-597X-9-33.

WHO. 2013. *Mental Health Action Plan 2013-2020.* Ginebra, Suiza.

WHO. 2018. “International Classification of Diseases (ICD-11).”

Zyoud, Sa’ed H., W. Stephen Waring, Samah W. Al-Jabi, and Waleed M. Sweileh. 2017. “Global Cocaine Intoxication Research Trends during 1975-2015: A Bibliometric Analysis of Web of Science Publications.” *SUBSTANCE ABUSE TREATMENT PREVENTION AND POLICY* 12. doi: 10.1186/s13011-017-0090-9.