Drug Abuse Research Trend Investigation with Text Mining

Li-Wei Chou 1st
Department of Physical Medicine and Rehabilitation
Cina Medical University Hospital
Taichung, Taiwan

Kang-Ming Chang 2nd
Department of Photonics and Communication Engineering
Asia University
Taichung 41354, Taiwan

Ira Puspitasari 3rd
Information System Study Program, Faculty of Science and Technology
Universitas Airlangga
Surabaya, Indonesia

***Abstract*—Drug abuse poses great physical and psychological harm to humans, thereby attracting scholarly attention. It often requires experience and time for a researcher,just entering this eld,to nd an appropriate method to study drug abuse issue.It is crucial for researchers to rapidly understand he existing research on a particular topic and be able to propose an eective new research method. Text mining analysis has been widely applied in recent years, and this study integrated the text mining method into a review of drug abuse research. Through searches for keywords related to the drug abuse, all related publications were identied and downloaded from PubMed. After removing the duplicate and incomplete literature, the retained data were imported for analysis through text mining. A total of 19,843 papers were analyzed, and the text mining technique was used to search for keyword and questionnaire types. The results showed the associations between these questionnaires, with the top ve being the Addiction Severity Index (16.44), the Quality of Life survey (5.01), the Beck Depression Inventory (3.24), the Addiction Research Center Inventory (2.81), and the Prole of Mood States (1.10). Specically, the Addiction Severity Index was most commonly used in combination with Quality of Life scales.Drug investigations have been carried out before by other researchers, but there is an analysis of drug abuse using the latest trends, namely by conducting association analysis.**

Keywords : Text mining , Drug abuse, addiction severity

I. Introduction

Because of the rapid development of information technology, information regarding various issues has been widely dispersed. Academia has long synthesized existing information and literature to acquire new knowledge by using a large amount of data. This is currently done by rst integrating analytical results from individual studies through systematic reviews and meta-analyses and then conducting statistical analyses to develop general conclusions. The present method may elicit discussions about the causal relationships and descriptions in studies as well as proposals of alternating treatment design to extend the implications of the literature. For example, a systematic review of drug literature in 2008 intended to determine the prevalence of illicit drug injection among people aged 15–64 years and the prevalence of HIV among injecting drug users [1]. A previous study reviewed 11,022 questionnaires to estimate the prevalence of illicit drug use in 61 countries. The obtained results revealed that 77 of the population worldwide aged 15–64 years used illicit drugs,with China,the United States, and Russia having the most users. In addition, the study indicated that approximately 3.0 million people (range 0.8–6.6 million people) worldwide who use illicit drugs might be HIV positive. Another meta-analysis explored the relationship between drug use and the high prevalence of skin and soft tissue infection. Data of 20 papers involving 9,502 patients presented a high correlation between the two [2]. In addition to improving statistical methods, advancements in information technology have facilitated the development of articial intelligence and big data algorithms, both of which have been extensively applied in Hindawi Computational and Mathematical Methods in Medicine Volume 2020, Article ID 1030815, 8 pages https://doi.org/10.1155/2020/1030815 various elds, particularly the elds of public health and biomedical information [3]. One of the numerous applications of big data is text mining.Based on natural language processing, this technique uses keyword matching and the connections between keywords to identify potentially useful information. Text mining has also been applied to biomedical research, rapidly extracting crucial information from a large amount of biomedical literature studies. Because automated screening makes reviews more ecient, numerous new tools have been introduced for text mining in biomedical research [4–6]. Nevertheless, this method is feasible only under the premise that researchers are procient in determining the usability,applicability,adaptability,

Identify applicable funding agency here. If none, delete this.

interoperability, and comparative accuracy of current text mining resources [7]. The existing research shows that text mining can reduce 30–70 of the workload of literature review [8]. Text mining has various applications. For example, to facilitate the development of precision medicine, text mining has been applied to examination of electronic medical records. The extensive use of electronic medical records provides clinicians and researchers with large amounts of data, which can be transferred to eective clinical care tools [9]. Another example text mining application is the use of narrative text analysis of electronic medical records to explore adverse drug reactions(ADRs) [10]. Researchers also applied text mining to clinical progress notes of cardiovascular diseases; text mining enabled them to calculate the probability of developing said diseases. The said study reviewed 282,569 echo cardiography report to identify patients with trileaetaortic stenosis (TAS) or coronary artery disease (CAD). The results revealed a positive predictive value of 0.95 compared with the standard of 0.53 by the International Classication of Diseases, Ninth Revision, Clinical Modication for TAS diagnosis and a positive predictive value of 0.97 compared with the standard of 0.86 for CAD diagnosis[11].ADRs of using aesthetic medicine,ranging from severe morbidity to mortality, indicate the importance of drug safety.In the past,the life cycle of a drug was monitored from drug development to clinical trials to detect safety problems at an early stage. The drug was continued to be monitored after marketing approval. The study also used text mining to identify potential safety concerns of drugs from source articles, including biomedical literature, articles posted by consumers on social media platforms,and narrative electronic medical records [12]. Applications of text mining can also be observed in drug and drug abuse research [13, 14]. For example, the study[15]developed a series of text mining procedures for designing new drugs. Using data from the DrugBank database, the said study aimed to determine how the chemical and protein compositions of drugs are related to disease-related genes and pathways to ultimately help develop new drugs[15].In another study,text mining was employed to explore the relationship between drug abuse and depression among young adults using 17,723 abstracts downloaded from PubMed.During the text mining process, keywords from these abstracts were organized, and a keyword cloud was used to present the topic content directly and demonstrate the term distribution for each topic. The results demonstrated that the association between drug abuse and depression among young adults lies in the links between keywords—such as sexual experience and violence—as well as risk factors of substance use among young adults. Text mining is also commonly employed in neurological drug abuse research [16]. The National Institute of Statistics and Censuses of Argentina investigated the prevalence of psychoactive substances in the country to estimate their consumption of psychoactive substances [17]. A study in the UK employed text mining and big data techniques to investigate the eectiveness of vareniclineas-apharmaceuticalaidforsmokingcessation. The aforementioned study employed association rule mining to analyze 46,685 individuals' data from the UK Health Improvement Network database. The results revealed that varenicline was most commonly prescribed to heavy smokers aged 31–60 years and those diagnosed with chronic obstructive pulmonary disease; varenicline was rarely prescribed to healthy people,people older than 60 years, light smokers, and smokers with mental illness or dementia [18]. Application of big data techniques to social networking data can also be used for drug abuse and addiction research [19]. A study examining the association between young adults and their non medical use of prescription medications analyzed 2,417,662 posts on Twitter. The said study discovered that 75.72 of tweets with URLs contained a hyperlink to an online aliate marketer that links directly to illegal online pharmacies where Valium can be bought without a prescription [20]. The aforementioned literature demonstrates that big data techniques in various forms have already been applied to academic research regarding drug abuse. This study applied text mining to organize drug abuse literature with the objective of understanding the current trends of drug abuse research using big data and association analysis. The results may serve as references for researchers to quickly understand large amounts of existing knowledge within their eld. **Text mining for investigating drug abuse using a database from the drugbank that aims to determine the chemical and protein position of the drug contained in the drug. To analyze the level of addiction of people to drugs can be done by bid data techniques in text mining.**

II. Materials and Methods

This study used the following keywords to search for and download drug abuse articles published till 2018 in PubMed: detoxication,addiction,drug abuse,substance,methadone, drug addiction, and therapy. End Note, bibliographic management software, was employed to organize the collected literature. A total of 28,488 articles were collected. After ltering out duplicate articles, those without an abstract (title, keyword, year, and author), non journal articles, and those with general terms (e.g., background, objectives, methods, results, conclusions, stop words, and numbers), 19,843 articles remained.The bibliographic data were stored in Excel les. Article data included the journal name,article title, abstract, keywords, authors, and year of publication. Dissertation data were analyzed using Poly Analyst (Megaputer Intelligence, Inc., Bloomington, IN, USA). The main Computational and Mathematical Methods in Medicine. **computing functions of Poly Analyst include data importing, data sorting, charting, classication, estimation, prediction, correlation,and clustering.The computing functions used in this study are text mining and link analyses [21]**. The text mining tool has capability for scalability, visual creation of analysis, interactive visualization, drill-down analysis, and execution of reports. It also includes automatic spelling correction, search for words and terms, detection of unpredicted issue,and adictionary editor for synonyms and abbreviations.

It has several steps for data analysis,which are as follows:

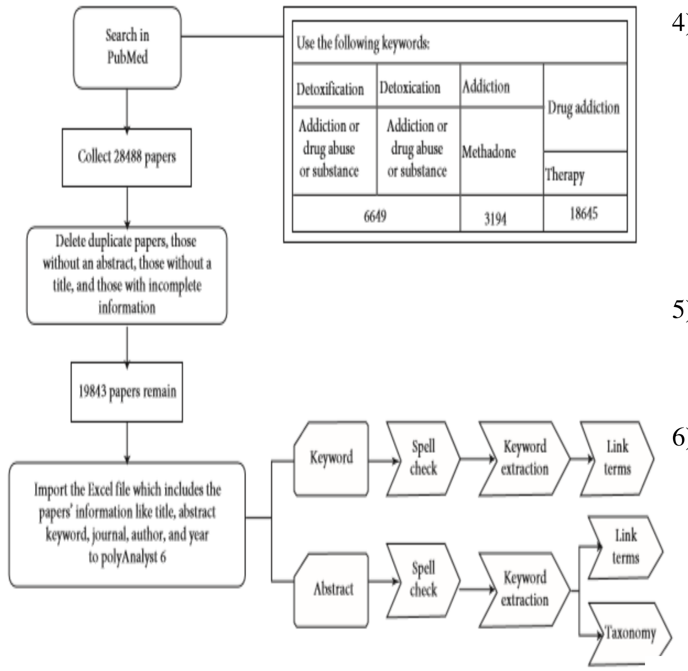

FIGURE 1: Experiment flowchart.

Fig. 1. flowchart

Use the following keywords:			
Detoxification	Detoxication	Addiction	Drug addiction
Addiction or drug abuse or substance	Addiction or drug abuse or substance	Methadone	Therapy
6649		3194	18645

Fig. 2. keywords cloud

1) Data loading: software process commands are written for text mining, functional nodes are connected to import the Excel les into Poly Analyst,and the parameter types of the data are adjusted.
2) Spell check: the spelling correction is conducted to improve the accuracy of the data content, thereby reducing the deviation of the data mining result from the actual situation. This procedure belongs to data cleaning, data transformation, and text segmentation.
3) Keyword extraction: this step comprises two tabs. The rst tab is for keyword extraction that comprises the investigated documents.It displays all records for a selected keyword with the word being highlighted. On the second tab,extraction is done to nd phrases and stable combinations of words.
4) Link terms: after completing the preprocessing task, keyword extraction and link analysis are conducted. A huge amount of correlated keywords and phrases is connected with a graph by a given connection tension threshold. As we modied the threshold, low-tension relations are hidden and the graph updates to only display the remaining links. By increasing the minimum tension threshold, we lter out a small number of records where there is relation between two words.
5) Creating taxonomy: the term "taxonomy" is generally dened as a classication system. In the taxonomy, all custom categories are created by users under neath the root category.
6) Visualizing the categorization result: during analysis, we can see some results in the taxonomy following visualization. The full analysis process to determine the distributions of academic drug research is illustrated in Figure1.The names of questionnaires commonly used in drug addiction treatment were extracted for text mining.

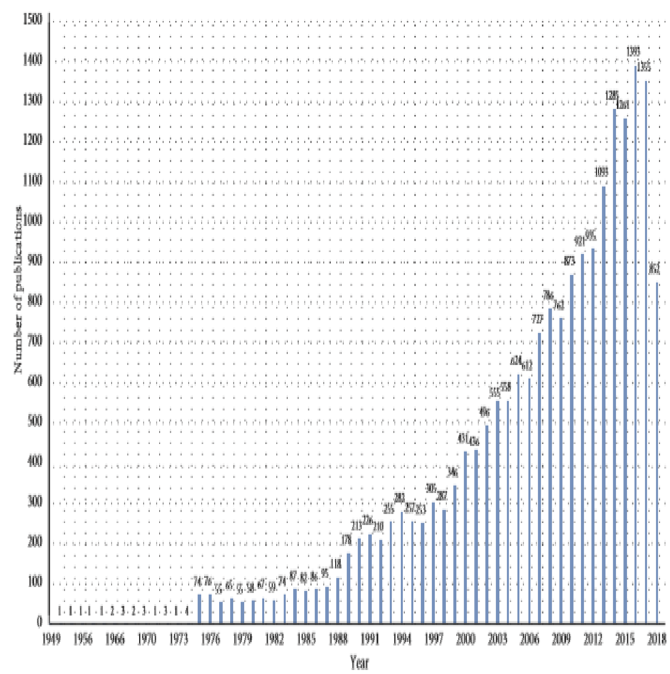

FIGURE 3: Publication distribution (by year).

Fig. 3. grafik

III. RESULT

The distribution of collected keywords was visualized with a keyword cloud (Figure 2). More frequently a keyword appeared,the larger the are a it occupied.In addition to the most frequently appearing keywords—treatment, study,addiction, drug, and patient—other terminologies related to drug addiction appeared. The numbers of dissertations in which these keywords appear are presented in Table 1. Arranging these publications by year result ed in the graph shown in Figure3.Since2013,the number of publications each year has exceeded 1,000, with the largest number of dissertations (1,393 papers) been published in 2016. Moreover,the number has increased with time,indicating that drug addiction treatment has received increasing attention from academia and suggesting the growth of future research. Analysis results revealed that, among all publications, 2,992 utilized questionnaires and scales. Sections discussing these measurement instruments were excerpted and organized.

Table 2: Questionnaire distribution.

	Questionnaire	Publication no.	Ratio (%)
1	Addiction Severity Index	492	16.44
2	Quality of Life	150	5.01
3	Beck Depression Inventory	97	3.24
4	Addiction Research Center Inventory	84	2.81
5	Profile of Mood States	33	1.10
6	Craving Questionnaire	23	0.77
7	Brief Symptom Inventory	21	0.70
8	General Health Questionnaire	18	0.60
9	Severity of Dependence Scale	18	0.60
10	Brief Pain Inventory	18	0.60
11	Minnesota Multiphasic Personality Inventory	15	0.50
12	Short Opiate Withdrawal Scale	13	0.43
13	Opiate Treatment Index	13	0.43
14	SF-36	12	0.40
15	Young Mania Rating Scale	11	0.37
16	Hospital Anxiety and Depression Scale	11	0.37
17	Pittsburgh Sleep Quality Index	11	0.37
18	Neuropsychiatric Inventory	11	0.37
19	Temperament and Character Inventory	10	0.33
20	State-Trait Anxiety Inventory	10	0.33
21	Mini International Neuropsychiatric Interview	10	0.33
22	Childhood Trauma Questionnaire	10	0.33

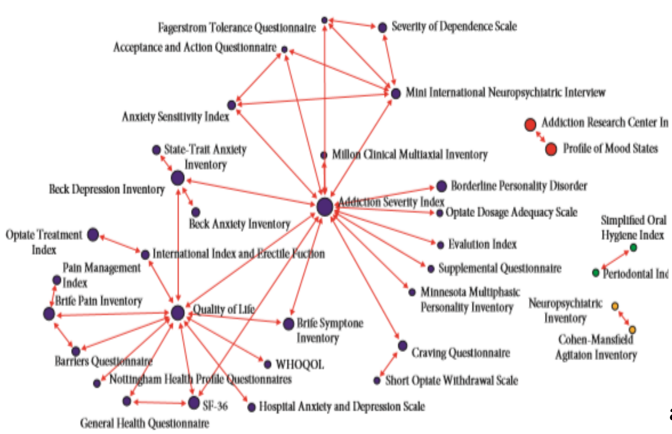

Fig. 4. Associations pattern between questionnaires

Questionnaires or scales used in more than 10 papers are presented in Table2, which shows that the most commonly employed assessment tools were the Addiction Severity Index (ASI, 16.44), Quality of Life (QoL, 5.01) scales, and the Beck Depression Inventory(BDI,3.24).Figure4 shows a diagram of the link analysis of the questionnaires. Visualization of the association analysis reveals that two clusters formed, with the ASI and QoL scales as the respective cluster centroids.

After further analysis of the association between questionnaires and the ASI as the cluster centroid, the most common questionnaires used in combination with the ASI were compiled into data shown in Table 3. Questionnaire combinations of the ASI with the QoL scale or the BDI were the most common assessment tools and there search direction most commonly approved by academia and clinical practitioners.For these condcluster with theQoL

Table 3: Commonest combinations of questionnaires.

	Questionnaire A	Questionnaire B	Tension	Support
1	Addiction Severity Index	Quality of Life	1.00	28
2	Addiction Severity Index	Beck Depression Inventory	0.99	22
3	Addiction Research Center Inventory	Profile of Mood States	0.88	19
4	Addiction Severity Index	Brief Symptom Inventory	0.67	11
5	Beck Depression Inventory	Quality of Life	0.26	11
6	Quality of Life	Brief Pain Inventory	0.39	7
7	Quality of Life	SF-36	0.33	7
8	Addiction Severity Index	Mini International Neuropsychiatric Interview	0.35	6
9	Quality of Life	General Health Questionnaire	0.35	5
10	Quality of Life	Brief Symptom Inventory	0.14	5
11	Addiction Severity Index	Craving Questionnaire	0.14	4
12	Addiction Severity Index	SF-36	0.11	4

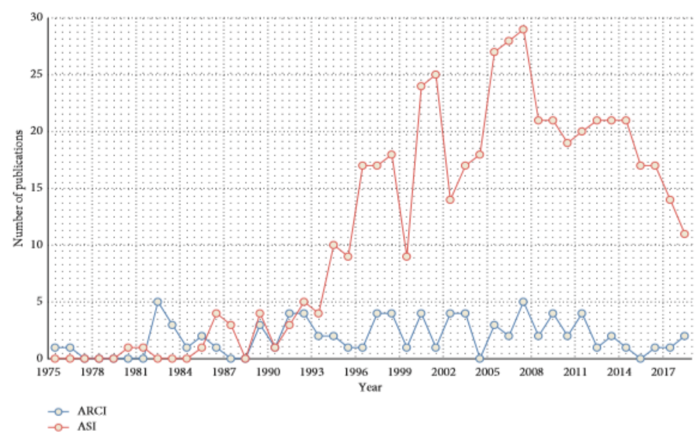

Fig. 5. Publications with use of the ASI/ARCI by year

scaleasthecentroid,someofthequestionnaireslinkedto itwere-alsolinkedtotheASIcluster,whereasotherswere evidently linked to only the QoL cluster, such as studies utilizing the Brief Pain Inventory, General Health Questionnaire, and Brief Symptom Inventory.

IV. Evaluation

Text mining cannot guarantee effective results, results from text mining may also not be effective. Preprocessing to delete excessive text before analysis is very important, but excessive preprocessing can also delete useful information, which will not be presented in subsequent analyzes. In addition, mining the text to organize references has many benefits, especially speed. Preprosessing is done using machine learning which is a java based open source package. After further analysis is carried out distributing

questionnaires to related people to be compiled into data. The assessment questionnaire adopted by most studies was ASI and Addiction Research Center Inventory (ARCI).

V. CONCLUSION

This study used text mining to explore the use of questionnaires in drug addiction research. The visualization techniques used with text mining enable researchers to rapidly determine how frequently each questionnaire type appears in all relevant research and the numbers of employed assessment tools by year. Future studies may leverage this method to select promising assessment tools to explore topics of their interest.

A. Acknowledgments

This research was funded by China Medical University Hospital, grant number CRS-108-047, and Asia University Hospital, grant number 10651006.

REFERENCES

[1] B.M. Mathers,L.Degenhardt,B. Philips et al ”Global epidemiology of injecting drug use and HIV among people who inject drugs” , *The lancet* (2008), 1733–1745.

[2] M. Moradi-joo, H.Ghiasvand, M.Noroozi et al ”Prevalence of skin and soft tissue infections and its related high-risk behaviours among people who inject drugs: a system review and meta-analysis”, *journal of substance* (2019), 1350-360.

[3] M.J. Khoury and J.P.A loannidis, ”Big data meets public health”, *science* (2008), 1054–1055

[4] C. Simon, K. Davidsen, C. Hansen, E.seymour, M.B.Barnkob, and L.R.Olsen Bioreader: a text mining tool for performing classification of biomedical literature, *BMC Bioinformatics* (2019), 57

[5] P.Y wu, C.W Cheng, C.D. Kaddi, J. Venugopalan,R. Hoffman, and M.D. wang, ”Omic and electronic helath record big data analytics for precision medicine”, *Transactions on Biomedical Engineering* (2017), 84–90

[6] M. Modaresnezhad, A. vahdati, H.Nemati, a. Ardestani, and F.sadri ”A rule-based semantic approach for data integration, standardization and dimensionality reduction utilizing the UMLS: application to predicting bariatric surgery out comes”, *computer biology and medicine* (2008), 84-90

[7] P. Przybyla, M.Shardlow, S.aubin, et al,, ”Text mining resources for the life science”, *Database* (2016)

[8] A.O’Mara-Eves,J. Thomas, J. MacNaught, M.Miwa,and S. Ananiodou, ”Using text mining for study identification in systematic review a systematic review pf current approaches”, *systematic review* (2015), 59

[9] M.Simmons, A.singhal, andZ.Lu, ”Text mining for precision medicine: bringing structure to EHRs and biomedical leterature to understand genes and health”, *Translational Biomedical Informatics* (2016), 139-166

[10] P.warrer, E.H. Hansen, L.Juhl-Jensen, and L.Aagaard, ”Using text mining techniques in electronic patient records to identify ADRs from medicine use”, *British Journal of clinical Pharmacology,* (2012), 674-684

[11] A.M. Small,D.H Kiss,Y.Zlatsin et al ”Text mining applied to electronic cardiovascular procedure reports to identify patients with trieaflet aortic stenosis and coronary artery disease”, *science* (2008), 1054–1055