Quantitative Analysis of China's New Drug Policy Based on Text Mining Perspective

Yufang He, Ruoyao Jia, Wenjing Duan, Zhen Gong, Chengye Zhang, Kaiyue Shen, and Fang Xia(✉)

School of Health Management, Changchun University of Chinese Medicine, Changchun 130117, China
Xiafang425@126.com

Abstract. This paper focuses on the vital domains and developmental trends of new drug-related policy. A total of 98 policy texts from 2010–2021 were subjected to cluster analysis and strategic coordinate analysis. The results showed that the number of new drug policies is fluctuating upward, the coordination of departments need to be strengthened; the emphasis of new drug policy is the innovative drug R&D for critical ailment, and the development direction is to deepen the reform of the new drug review and approval system. This study provides insight into, and a resource for the further exploration of New Drug Policy.

Keywords: New drugs · policy analysis · text mining

1 Introduction

The national policies and reform measures are critical in fostering the growth of the new drug industry [1]. Not only do new drug policies have a direct impact on the Chinese people's drug rights and health, but they also influence the pharmaceutical industry's development direction to a degree. Chinese government and related departments have guided by clinical value and the drug review and approval system as a springboard to develop stringent laws, regulations, and industry standards for the development of China's new drug industry. This encompasses the entire process of developing, registering, manufacturing, operation and other aspects. Research on new drug-related policies has also attracted widespread attention from domestic and foreign experts and scholars. The majority of studies have examined policy measures and their effectiveness at specific levels, either macro or domain-specific, and only a few have conducted quantitative analyses of the content and topics of new drug-related policy texts. This article employs a mining perspective to conduct a quantitative analysis of 98 national-level new drug policy texts that have been collected and collated.

The purpose of this study is to examine the external characteristics and thematic characteristics of new drug-related policies, to identify key points and development trends in the implementation of new drug-related policies, and to serve as a resource for future policy improvement.

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 750–759, 2023.
https://doi.org/10.2991/978-94-6463-198-2_78

2 Date and Analysis

2.1 Data Sources

Data for this study were collected from the State Council, the Development and Reform Commission, the Ministry of Science and Technology, the Health and Welfare Commission, the State Intellectual Property Office, the State Drug Administration, the Taxation Bureau, the Ministry of Industry and Information Technology and other official state websites. The text was screened and selected according to the following principles to ensure the accuracy of the selection of information: ① Data for this study were collected from 2010 through 2021. ② A total of 98 policy texts on "new drugs", "new drug research and development", "drug development", etc. were published.

2.2 Data Analysis

Data from this study were analyzed using ROSTCM software, and the policy research topics were identified by combining the valid keywords obtained from the screening. The co-occurrence analysis of effective keywords aims to find out the correlation between the subject words in the policy text and sort out the policy context.

3 Research Results

3.1 Analysis of Posting Time

Our preliminary data showed a trend of the number of policy promulgations showed a fluctuating upward trend, with an annual average of 9.8 policy promulgations and a maximum of 16 (Fig. 1).

3.2 Analysis of the Issuing Agency

It is found that there are 10 issuing departments involved in the statistics of policy issuers, of which NMPA is the most important, with a total of 28 papers issued between 2010 and 2021, accounting for 28.6% of the total sample, followed by the State Council, the National Development and Reform Commission, and the Intellectual Property Office (Table 1). Calculating the joint texts, we found that the number of joint texts issued by the Ministry of Finance, the State Administration of Taxation and the Ministry of Science and Technology is relatively high, the departments with a relatively high number of separate texts are the State Council and NMPA. The total number of jointly issued texts is 26, accounting for only 18.4% of the total number of texts.

3.3 Word Frequency Analysis

Word frequency analysis is one of the important means of text content analysis; the higher the number of occurrences of a word, the more critical the word is in the text [2]. In order to ensure the rigor of keyword extraction, some words with similar expressions were combined when counting the word frequency. Words with low relevance to the

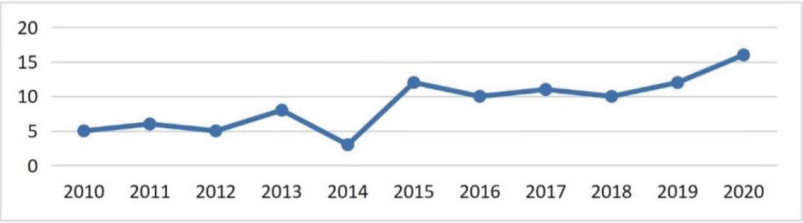

Fig. 1. Statistics on the time of issuance of new drug policies

Table 1. Agencies that publish new drug policies

Issuing Department	Number
NMPA	42 (10)
State Council	23
Development and Reform Commission	21 (8)
Intellectual Property Office	15 (5)
Ministry of Science and Technology	13 (8)
Ministry of Finance	12 (5)
Health Care Commission	8 (3)
State Administration of Taxation	7 (3)
Ministry of Industry and Information Technology	5 (3)

research topic were also included in the filter word list to improve the accuracy of data screening. After further extraction, the word frequency table that matched the scope of the study topic was obtained as the valid keyword word frequency table for this study (Table 2).

3.4 Co-term Analysis

ROSTCM was also employed to process the co-occurrence of the subject keywords and generate a valid co-word matrix after the generated valid keywords were collated (Table 3). The data on the diagonal line represent the co-occurrence frequency of the keyword in all policy texts, and the data on the non-diagonal line represent the co-occurrence frequency of two identical keywords in the same policy text. The higher value of the joint frequency of two keywords represents the stronger connection of the policy-focused topic [3].

3.5 Cluster Analysis

Cluster analysis techniques are a commonly used multivariate statistical analysis method in classification problems, also known as cluster analysis [4]. This study uses the "Dispersion Sum of Squares" clustering method, selects "Euclidean Distance" as the criterion

for interval division, and outputs as the result of the cluster dendrogram according to the variable z score (Fig. 2). Cluster dendrograms can map the clustering process more accurately, specifically to the extent in which individual samples are merged during the clustering process [5]. Calculate the average number of keyword co-occurrences within a class group as the density of the topic class, and the average number of co-occurrences of a topic with other subject groups as the centripetality [6]. It can be calculated that the spatial centripetality and density of various thematic groups are as shown in Table 4.

In this paper, the cluster distance was set at 10, which divided all keywords into five subject clusters, numbered with the letters A, B, C, D, and E. Keywords such as "pharmaceutical production", "enterprise", "holder" and "listing" jointly form theme group A, keywords such as "declaration", "evaluation" and "registration" jointly form theme group B, and "innovation", "research and development" and "drug" are combined into theme group C through high interrelatedness. The cluster distance between "quality" and "clinical trial" is relatively loosely correlated with other studies, so they are divided into two subject groups D and E separately.

3.6 Strategic Analysis

Strategic coordinate analysis is a common word analysis method that analyzes the evolution of various topics and the development process based on the position of the topic. In the strategic coordinate chart of this study, the X axis represents the maturity of the theme group in the policy provisions, and the Y axis represents the core degree of the theme group in the policy provisions. The density of the object group calculation is used as the abscissa, the centripetal degree is the ordinate, and the intersection point of the density and the average centripetal degree is the origin of the strategic coordinate diagram in Excel (Fig. 3).

Core areas. Thematic clusters located in the first quadrant are characterized by high density and high centripetality. The key words within the group are not only closely related, but also have a strong connection with other thematic groups. Thematic clusters in this quadrant tend to represent core areas of industry research. As can be seen from Fig. 3, the current core keyword theme group of new drug-related policies is Theme C, which shows that the core research and policy support system in the field of new drug policy is mainly the research and development of innovative drugs for major diseases.

Future trends. Thematic clusters in the second quadrant are characterized by low density and high centripetality. The research direction of the group in this quadrant represents the future development trend of the industry. The group at this quadrant is Theme B, which represents the focus of new drug policy formulation, indicating that new drug-related policies will continue to deepen around the theme of new drug review and approval in the future.

Marginal fields. Thematic clusters located in the third quadrant are characterized by low density and low centripetality. The exploration of the research theme of the group in this quadrant is still in the shallow stage of new drug policy formulation. This type of thematic group mainly focuses on the new drug marketing authorization holder system, and the focus is mainly on the drug modernization management system represented by the holder.

Table 2. Top 20 of the key word

Serial number	Keyword	Word frequency
1	Quality	491
2	Clinical Trials	410
3	Innovation	350
4	Registration	310
5	Filing	290
6	Review	270
7	New Drug	256
8	Applicant	225
9	Drug Administration	210
10	Drug	198
11	Listing	181
12	R& D	170
13	Holders	167
14	Pharmaceutical	152
15	Manufacturing Classification Management	145
16	Enterprise	138
17	Pharmaceutical business	125
18	Medical institutions	117
19	Adverse Reactions	106
20	Disease	95

Note Number of joint issuances in parentheses

Independent research field. Thematic clusters in the fourth quadrant are characterized by low centripetality and high density. The policy provisions represented by this type of delegation have been relatively perfected, but cooperation with other research directions needs to be further strengthened. The two thematic clusters D and E in this quadrant both conduct business research around their respective central themes, primarily drug quality management and evidence-based clinical trial research centered on drug safety.

4 Discussion

4.1 The Number of Formulation

The number of new drug policies formulated is generally on the rise and has always played an important role in the national science and technology strategy. With the innovation of modern technology and the development of national economy, the government's

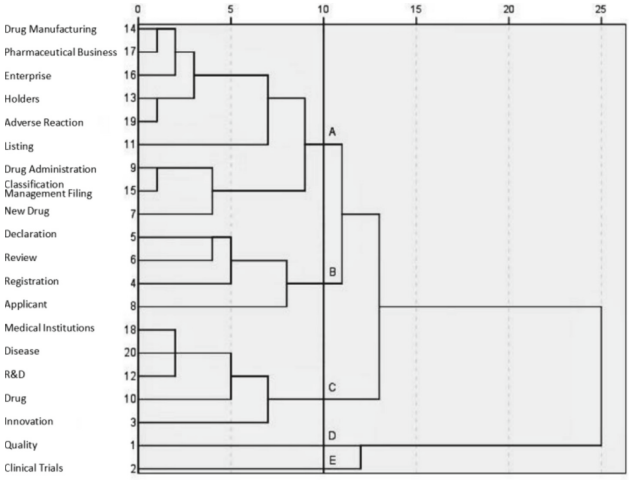

Fig. 2. Keyword clustering

attention to new drugs has been increasing. The study discovered that new drug R&D-related policy releases have obvious phase characteristics, such as a period of intensive policy introduction, and a period of time may be less. As a long-term system project, the stability and continuity of new drug policies are particularly important to the new drug industry.

4.2 Issuing Units

There are more agencies issuing new drug-related policies, involving more than 10 agencies and organizations. This is because new drug policies involve various aspects of new drug R&D, new drug registration, new drug marketing, and industrial management, and it is difficult to establish a sound, scientific, and effective policy system by the power of a single institution [1]. The departments that formulate new drug R&D-related policies alone are mainly 10. Only 26 of the 98 policy texts were issued jointly, accounting for 18.4% of the total, which indicates that the coordination and coordination ability of the policy-making departments needs to be improved.

4.3 Key Areas

As the central driver of new drug development, "innovative drug development for major diseases" has been an important area of policy development in China. Among them, innovative drug legislation and policy reforms for the prevention and treatment of major diseases have been frequently implemented. In 2006, the State Council, in accordance with the "National Medium and Long-term Scientific and Technological Development Plan (2006–2020)", deployed the "Major New drug Creation" national science and technology major special project to deploy industrial R&D chains around major diseases that endanger human health, with significant social benefits. In 2015, "Made in China 2025"

Table 3. Partial list of co-word matrices

keyword	1	2	3	4	5	6	…	19	20
Quality	491	211	206	156	108	127	…	108	199
Clinical Trials	211	410	262	113	123	98	…	106	204
Innovation	206	262	350	86	134	90	…	0	158
Registration	156	113	86	310	121	109	…	0	8
Filing	108	123	134	121	290	124	…	0	0
Review	127	98	90	109	124	270	…	0	0
New Drug	14	113	0	39	36	11	…	0	0
Applicant	56	0	0	98	46	45	…	0	0
Drug Administration	8	10	29	43	0	0	…	0	0
Drug	208	217	154	0	0	0	…	58	45
Listing	185	99	45	32	54	9	…	32	0
R&D	156	310	141	35	41	0	…	0	145
Holders	87	92	42	12	47	25	…	116	0
Drug manufacturing	68	4	0	0	0	0	…	126	0
Classification Management	11	11	58	46	0	13	…	0	0
Enterprise	71	5	0	4	9	13	…	0	0
Pharmaceutical business	80	0	0	0	0	0	…	125	0
Medical institutions	202	201	110	13	96	0	…	6	152
Adverse Reactions	108	106	0	0	0	0	…	106	2
Disease	199	204	158	8	0	0	…	2	95

Table 4. Keyword Topic word clustering results

Class group	Name of Theme Class Group	Centripetal degree	Density
A	New Drug Launch Management	434.75	271.50
B	New drug registration and approval	212.78	352.00
C	Innovative drug development	617.60	431.20
D	Quality	2261.00	0.00
E	Clinical Trials	2179.00	0.00

mentioned the development of new products of chemical drugs, traditional Chinese drug and biotechnology drugs for major diseases, and innovative drugs for the prevention

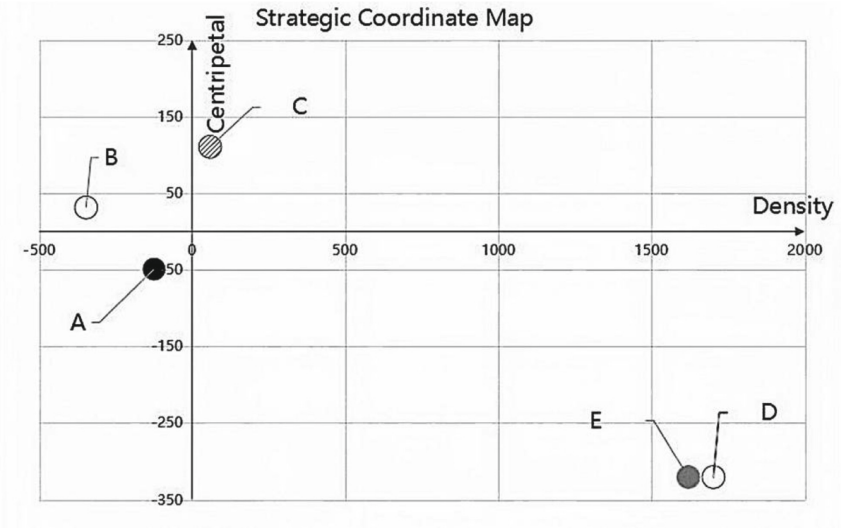

Fig. 3. Cluster grouping results strategy coordinate map

and treatment of major diseases have attracted widespread attention from the industry. In 2017, the transformation and upgrading of China's drug industry from "generic" to "innovative" was further promoted by the "General Administration's Opinions on Encouraging Drug Innovation and Implementing Priority Review and Approval", which proposed to accelerate the review process of drugs with obvious clinical value and major disease prevention and treatment. China accelerated a series of initiatives aimed at lowering the price of generic drugs to force drug companies to transform innovative drug research and development while encouraging innovation, such as generic drug consistency evaluation, quantity-based procurement, etc. The research and development of innovative drugs for the prevention and treatment of major diseases is not only related to public health rights and interests, but also to the international competitiveness of China's pharmaceutical industry, which has become the key focus of new drug policy formulation.

4.4 Development Direction

The "New drug review and approval system" is an important development direction for the introduction and reform of China's new drug policy. New drug development is a gradual process. The government has introduced a series of new drug registration and approval policies and reform measures to adjust the direction of research and development, enhance new drug research and development capabilities, and promote the long-term development of the pharmaceutical research and development industry. Since 2015, the priority review and approval system has been gradually established in China to encourage the R&D of innovative drugs that meet clinical needs. In 2017, the State Council proposed that the drug review and approval system should be combined with innovative drug research and development, intellectual property rights and drug

patents to build an innovative national supporting policy from the "Opinions on Deepening the Reform of the Review and Approval System to Encourage Drug and Medical Device Innovation", which is a milestone for the creation of new drugs in China. The new "drug Registration and Administration Measures" will be formally implemented in 2020, combining clinical trials of new drugs, drug listing licensee system and other in-depth reforms to ensure the safety, effectiveness and quality control of drugs to lay the foundation of the legal system. The reform of China's new drug policy will continue to focus on the review and approval system, and gradually cover various aspects such as special funds for new drug research and development, clinical trials, market circulation, intellectual property protection, drug marketing license holder system, and the inclusion of innovative drugs in the medical insurance catalogue. The coordination and implementation of other related policies for new drugs is an important development direction of the new drug review and approval system.

5 Conclusion

First, the overall increase in the number of new drug policy formulations, and the government's attention to new drugs is increasing. Second, the coordination and coordination ability of policy-making departments need to be improved. Finally, focus on innovation for major diseases. Drug research and development is a hot spot in the formulation of new drug policies, and the new drug review and approval system is coordinated with other relevant systems. For the development direction of future new drug policy.

References

1. Fan Yulu, Chen Yuwen. "Textual analysis of policies related to new drug R& D in China from the Ninth Five- Year Plan to the Twelfth Five-Year Plan [J]. China Journal of New Drugs, 2017, 26(13):1488- 1493.
2. Bao Pengfei, Meng Zhuo, Lu Xinyue, Wang Hongyun, Zheng Qiuying, Kong Junhui, Xu Kun, Li Ruifeng, Huang Youliang. Quantitative analysis of China' s Internet healthcare policy based on text mining perspective[J]. China Hospital,2021,25(10): 50-53.DOI:10. 19660/j.issn. 1671-0592.2021. 10. 14.
3. Zhang YH, Huang WD. Analysis of the progress of China' s online public opinion research (2014 - 2018) - based on strategic coordinate chart [J]. Journal of Nanjing University of Posts and Telecommunications (Social Science Edition), 2019, 21(03): 31-43. DOI:https://doi.org/10.14132/j.cnki.nysk.2019.03.004.
4. Feng Chujian, Xie Qijun. A comparative study on the performance of domestic and foreign PV industry policies [J]. China Science and Technology Forum, 2017(02) :58-65. DOI:10. 13580/j.cnki.fstc.2017 .02 .009.
5. Zhang Q. A review of the application of word frequency analysis method in the study of disciplinary development dynamics[J]. Library Intelligence Knowledge, 2011(02): 95–98+ 128. DOI:10. 13366/j.dik.2011.02.021.
6. Li Hanko, He Yang. A study on the attention to quality of government information disclosure in China- - an analysis based on policy texts[J]. Library,2021(08) :25-32.