

Science Popularization of Cancer under the New Mode of Internet-based Health Communication

Zijian Wu¹, Yeying Song^{1,*}

¹*Affiliated Cancer Hospital & Institute of Guangzhou Medical University, No.78, Hengzhigang Road, Yuexiu District, Guangzhou, 510095, China*

**Corresponding author*

Keywords: Cancer, prevention and treatment, science popularization, WeChat, hot topics

Abstract: Objective: This article is designed to analyze the health issues related to cancer specialties that are most concerned by the public and guide cancer medical institutions to actively take advantage of new communication modes such as the Internet to promote health science popularization. Method: We obtained the information pushed by the WeChat official accounts of eight grade A tertiary cancer hospitals in China from January 1, 2020 to December 31, 2021 through Python, screened the information related to health science popularization of cancer, and evaluated the hot topics and demands of the public on science popularization of cancer through the indexes of Reads, Favorites, and Likes. Result: The top five science popularization keywords are treatment, cancer, eating, food, and surgery; the top five diseases are: lung cancer, breast cancer, gastric cancer, nasopharyngeal cancer, and liver cancer; after text clustering, it is found that there are also plenty of messages themed healthy diet and postoperative metastasis and treatment. Conclusion: Internet-based communication platforms such as WeChat official accounts have become the mainstream approach of health science popularization and deserve attention and analysis. The results and methods of this study have guiding significance for hospitals to assess the cognitive needs of the public and hot spots of science popularization concerns in a timely manner, and can facilitate continuous improvement of science popularization.

Cancer has become one of the principal diseases affecting human health nowadays. According to a report released by International Agency for Research on Cancer of the World Health Organization, the number of new cancer cases and death cases in China were ranked the first in the world in 2020. In fact, the incidence and mortality rate of oncological diseases in China have been increasing progressively year by year. According to the report, the annual average growths of cancer incidence rate and cancer mortality rate in China reach 3.9% and 2.5% respectively over the past 15 years. It indicates that cancer has posed a great danger to the life and health of the Chinese people, and the situation of prevention and control is extremely grim^[1]. According to the guiding opinions of the World Health Organization, “one third of cancers are completely preventable, one third of cancers can be cured through early detection, and one third cancers can be treated with existing medical measures to prolong life, reduce pain and improve quality of life”^[2]. Therefore, the early discovery and early prevention of cancer have become the focus of national health care development. To be specific, health science popularization plays an indispensable role in cancer prevention and

treatment^[3]. With the rapid development of mobile Internet technology, media such as WeChat official account provides a new mode for health science popularization. Many hospitals have achieved desirable results in health science popularization through the Internet^[4]. This study uses text mining technology and language analysis technology to mine and analyze the science popularization messages pushed by the WeChat official accounts of grade A tertiary cancer hospitals in China to understand the focus and concerns of the public on cancer, so as to provide reference for health science popularization of cancer.

1. Data and Methods

1.1 Research Sample

In this study, eight grade A tertiary cancer hospitals in China (including Peking University Cancer Hospital, Fudan University Shanghai Cancer Center, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Jiangsu Cancer Hospital, Sichuan Cancer Hospital, Tianjin Medical University Cancer Institute & Hospital, Zhejiang Cancer Hospital, China Medical University Cancer Hospital, and Cancer Hospital of Chinese Academy of Medical Sciences) were selected as the research samples, and the messages related to science popularization and health pushed by their WeChat official accounts from January 1, 2020 to December 31, 2021 were acquired for statistical analysis. Among the above eight hospitals, Affiliated Cancer Hospital and Institute of Guangzhou Medical University is a grade A tertiary cancer hospital affiliated with a “Double First-Class” university, and the oncology specialty is a high-level key specialty of Guangdong Province. The other seven hospitals have been selected as the top ten in the National Oncology Specialty Ranking issued by the Hospital Management Institute of Fudan University.

1.2 Data Acquisition and Processing

In this study, we used Python 3.7 to obtain the tweeting data of the eight WeChat official accounts using a crawler program, including the name of the official account, article title, article content, release time, number of reads, number of likes, number of wows, and URL address. Statistics showed that there were a total of 546 articles related to health science popularization. According to the evaluation indexes used in the “National Science Popularization We-Media (WeChat) Top List”^[5], the weight of the index of reads is 90%, and the weight of the index of likes is 10%. The total evaluation score of each science popularization article was obtained by summing up the scores of reads and likes, and 405 articles scoring over 1000 were selected. The textual contents of these articles were included in the corpus of health science popularization texts.

1.3 Research Methods

(1) Extraction of keywords of health science popularization

In this study, we used Jieba, a Chinese word segmentation tool, to accurately segment the content of the articles. The Chinese stop-words list, HIT stop-words list, and Baidu stop-words list as well as the noise-words list extracted by the author after analysis were used to screen and filter the words or symbols that do not have a strong relationship with the emotional information or thematic information contained in the text, hoping to facilitate the analysis of the hot topics in the corpus of cancer science popularization information.

(2) Extraction of text features

Term frequency-inverse document frequency (TF-IDF) is a statistical method often used to extract text features, which is in fact $TF * IDF$. TF refers to the frequency of a given term in the

document, and IDF is a measure of the general importance of a term. The IDF of a particular term can be obtained by dividing the total number of documents by the number of documents containing the word, and then getting the logarithm to the base 10 of the quotient.

(3) Text clustering of health science popularization

First, the initial text clustering was performed by K-means. The TF-IDF text features further extracted from the health science popularization text information were automatically aggregated into k initial cluster centroids (k=10 in this study) by unsupervised learning, and each cluster was defined as a category and summarized. If there are similar categories in the classification results, these categories would be further summarized based on the clustering results to obtain the final clustering results, so as to identify which categories of health science popularization contents are more concerned by users.

2. Statistical results

2.1 Statistical results of health science popularization keywords

We carried out a statistical analysis of the key words extracted from the text corpus of health science popularization topics and found that the top 20 key words were: treatment, cancer, eating, food, surgery, examination, lung cancer, nutrition, disease, breast cancer, clinical, risk, radiotherapy, symptom, chemotherapy, diet, prevention, drug, prevention and treatment, and breast. Through observation, we can find that cancer and treatment have the highest frequency among the top 20 keywords. Therefore, we conducted further word frequency statistics for specific cancer types, and the most frequent cancers included lung cancer, breast cancer, gastric cancer, nasopharyngeal cancer, liver cancer, prostate cancer, and esophageal cancer.

2.2 Clustering results of health science popularization topics

Table 1: Detailed Information of Text Clustering

Category	Number of Articles	Key Features	Content Examples (Article Title)
Dietary risk awareness promotion	89	Eating, food, disease, diet, risk	Drinking 100 ml of sugary drinks a day can increase the risk of cancer by as much as 18 percent.
Postoperative metastasis and treatment	88	Treatment, surgery, metastasis, chemotherapy, drugs	This body part is no longer a “surgical restricted area”.
Lung cancer and related knowledge	45	Lung cancer, smoking, Chinese medicine, disease, symptoms	There is more than one cause of lung cancer, but why do we emphasize so much on quitting smoking?
Gastric cancer and related knowledge	44	Gastric cancer, examination, pylorus, Helicobacter, esophageal cancer	Who needs a gastrointestinal endoscopy? Is it scary to have a gastrointestinal endoscopy?
Anti-cancer psychological counseling	39	Cancer, anti-cancer, treatment, chemotherapy, psychology	@ All family members, are you really ready to accompany your cancer patients?
Cancer prevention and treatment knowledge dissemination	31	Prevention and treatment, cancer, cancer prevention, dissemination, knowledge	Some people are more likely to get esophageal cancer than others! Who are these people?
Scientific and nutritious diet	26	Food, eating, diet, meals, intake	What should cancer patients eat? Nutrition experts will tell you.
Gynecologic cancer-related knowledge	25	Breast, breast cancer, women, treatment, screening	How should breast cancer screening be done?
The importance of physical exercise	18	Exercise, weight, walking, obesity, body	Can obesity also induce cancer?

Through analysis of the text clustering of the corpus of health science popularization, we finally obtained nine categories of topics that were widely concerned by users. The nine categories of topics were further classified into four categories (see Table 1 for details). The number of articles

pushed in each category showed that the number of articles in the categories of diet and health and postoperative metastasis and treatment was relatively high, which also represented the hot topics concerned by the public.

3. Discussion

Relevant research shows that cancer is closely related to behavioral habits, weight, diet, and life style, and effective control of various adverse triggers can reduce the incidence of tumor diseases to a certain extent. In addition, the technology of early diagnosis and early screening of tumor has become increasingly more mature, and the early discovery of cancer will greatly improve its cure rate. However, as there still lacks clear and effective treatment for cancer at present, strengthening the science popularization of cancer prevention and treatment, improving people's understanding of cancer and awareness of disease prevention, and regulating healthy behaviors and habits will play a positive role in reducing the harm of cancer and improving the survival time and quality of life of cancer patients. After the new media such as WeChat official account has become the mainstream means of health science popularization, the information of cancer prevention and treatment has been disseminated more through Internet mobile media. Therefore, the contents of science popularization information pushed by medical institutions, especially the WeChat official accounts of the cancer hospitals as well as the public attention to the information deserve our analysis^[5].

Through the results of this study, we found that the diseases with the highest frequency in the popularization information pushed by the eight cancer hospitals, except for nasopharyngeal carcinoma, were all among the top ten cancers in the 2018 China Cancer Registry Annual Report, which indicates that the contents of the popularization information pushed by the hospitals and the cancers that people pay attention to are basically in line with the trend of epidemiology. In the top 20 keywords, except for specific diseases, most of them were about two main categories of "prevention" and "treatment", which indicates that the focus of the public on cancer science popularization has been more clearly positioned to prevention and treatment. In addition, the analysis of the results of health science popularization topics shows that cancer prevention and treatment and rehabilitation are the main hot issues of concern, which indicates that the public's knowledge of cancer has been extended to the perspective of cancer prevention and rehabilitation.

4. Conclusion

Based on the analysis of the results, we can find that the public's cognitive demands for cancer science popularization information has become increasingly strong, and their perspective of focus has also become more comprehensive and scientific. At present, the contents of science popularization promoted by grade A tertiary cancer hospitals in China are basically in line with the strategic policies of cancer prevention and treatment. However, with the increasing development of information technology, Internet platforms such as WeChat provide more diversified ways for the public to obtain medical information. Therefore, hospitals should make active use of the Internet and other information technology means to make sure that the contents of science popularization can not only reflect the up-to-date medical advances, but also meet the cognitive needs of the public.

Acknowledgement

Research Fund: This paper is funded by the 2020 Joint Project of the "Fourteenth Five-Year Plan" for the Development of Philosophy and Social Sciences in Guangzhou (Project No. 2022GZGJ68) and the 2022 Characteristic Innovative Projects in Regular Institutions of Higher

Learning in Guangdong (Humanities and Social Sciences) by Guangdong Provincial Department of Education (Project No. 2022WTSCX088).

References

- [1] Ning, X. Y. (2018). *LinkDoc: Delivering the value of medical big data*. *Entrepreneur*, (9), 3.
- [2] Yang, L. (2019). *There is no need to be jittery at the mention of cancer*. *China Today*, 68(04), 80.
- [3] Li, J. B., & Li, N. (2021). *Three lines of defense to fight back cancer*. *Well-being*, (12), 2.
- [4] Zhang, X. D., Gao, F., Jia, L. T., Wang, W., Qu, Y. Q., & Zhang, P. (2021). *Exploration and practice of innovative development path of cancer health science popularization-take the National Cancer Prevention and Treatment Science Contest held by the National Cancer Center as an example*. *Medical Education Management*, 7(03), 345-350.
- [5] Feng, Y., Fang, N., & Zhang, Q. X. (2018). *Modern Hospital*, 18(02), 220-223.