



Text analysis of Macao's COVID-19 prevention and control policies: discussion on strategy evolution and public health capabilities

Kitcheng Wong¹, Boyuan Wang², Kaichin Hsieh³, Chitin Hon^{4,5,6,7}, Zhiqi Zeng⁶, Laisan Ngai⁸, Zizheng Liu⁹, Yangqianxi Wang⁵, Sengtong Hon¹⁰, Honsam Lao⁸, Guibin Lu⁴

¹Constitutional Law and Administrative Law, School of Law, Wuhan University, Wuhan, China; ²School of Computer Science and Engineering, Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China; ³UCL Faculty of Engineering Sciences, University College of London, London, UK; ⁴Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China; ⁵Guangzhou Laboratory, Guangzhou, China; ⁶Respiratory Disease AI Laboratory on Epidemic Intelligence and Medical Big Data Instrument Applications, Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China; ⁷Guangzhou Key Laboratory for Clinical Rapid Diagnosis and Early Warning of Infectious Diseases, KingMed School of Laboratory Medicine, Guangzhou Medical University, Guangzhou, China; ⁸Kao Yip Middle School, Macau, China; ⁹Faculty of Law, Macau University of Science and Technology, Macau, China; ¹⁰School of History and Culture, South China Normal University, Guangzhou, China

Contributions: (I) Conception and design: C Hon, K Wong, G Lu, Z Zeng; (II) Administrative support: K Wong, B Wang, K Hsieh, L Ngai; (III) Provision of study materials or patients: C Hon, Z Liu, Y Wang, S Hon; (IV) Collection and assembly of data: G Lu, H Lao, S Hon, L Ngai; (V) Data analysis and interpretation: K Wong, Z Liu, B Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Guibin Lu, MD. Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, China. Email: gblu@must.edu.mo.

Background: The global impact of the coronavirus disease 2019 (COVID-19) pandemic has been profound. Macao Special Administrative Region (SAR), renowned as an international hub for tourism and entertainment, has actively responded to the crisis. However, a comprehensive analysis detailing the evolution of Macao SAR's policies throughout this period is currently lacking.

Methods: This study aims to comprehensively understand the decision-making processes, policy formulation, and implementation strategies of the Macao SAR government amidst the pandemic through the analysis of speeches and inquiries made by legislative council members and other relevant documents. Employing both quantitative and qualitative analytical methods, including word frequency analysis and word vector models, we identify key themes and patterns. Additionally, we conducted a comparative analysis of keyword frequencies during the two waves of the pandemic using radar charts.

Results: The results indicate a heightened focus by the Macao SAR government on pandemic control measures and economic impacts. In response, the government formulated and implemented policies, provided support initiatives, and managed port clearance, all while focusing on enhancing healthcare infrastructure and community services.

Conclusions: The government persistently amends its policies in response to the evolving challenges posed by the pandemic. The evolution of the dynamic Zero-COVID strategy highlights the government's adaptability and comprehensive consideration, ensuring public health and societal stability.

Keywords: Macao Special Administrative Region (Macao SAR); coronavirus disease 2019 pandemic (COVID-19 pandemic); public health strategies; natural language processing; Word2Vec

Submitted Nov 29, 2023. Accepted for publication Jan 11, 2024. Published online Jan 29, 2024.

doi: 10.21037/jtd-23-1818

View this article at: <https://dx.doi.org/10.21037/jtd-23-1818>

Introduction

Background

Coronavirus disease 2019 (COVID-19) is a highly contagious respiratory disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (1). The virus spreads rapidly through droplet transmission, facilitating swift human-to-human transmission. With its prolonged incubation period and rapid transmission rate, the urgency and complexity of epidemic prevention and control have been heightened. The global outbreak of the COVID-19 pandemic has had profound effects on societies worldwide. Since its initial emergence at the end of 2019, the virus has swiftly disseminated across the globe, resulting in millions of infections and hundreds of thousands of fatalities (2). The global health crisis has posed significant challenges, impacting not only public health systems but also various sectors, such as the economy, education, and tourism. In response to the pandemic, governments worldwide have implemented extensive measures to prevent and control the spread of the virus. These measures encompass city lockdowns, restrictions on mobility, the enforcement of social distancing, and the reinforcement

of hygiene and protective measures. Concurrently, governments have enacted a series of emergency policies and initiatives to support affected businesses and individuals, ensuring the fulfillment of basic life necessities for the public (3).

As a Special Administrative Region (SAR) of China, the Macao SAR has taken proactive measures in response to the COVID-19 pandemic (4-6). Situated in the Pearl River Delta region adjacent to Hong Kong, Macao SAR is an international hub for tourism and entertainment. Its unique geographical location and reliance on the tourism industry make it particularly susceptible to the rapid spread of infectious diseases. Consequently, the government's response to the pandemic is crucial for safeguarding public health, ensuring resident well-being, and maintaining social stability. This demonstrates a profound comprehension of specific preventive measures, highlighting flexibility and diversified thinking in response to sustained pandemic pressures (7-9).

During the pandemic, the Macao SAR government swiftly implemented a series of preventive and control measures. Firstly, they strengthened border controls, restricted entry, and enforced rigorous quarantine and isolation measures. Secondly, they implemented widespread social distancing measures, including the closure of entertainment venues and the cancellation of large gatherings. Additionally, the government intensified public health campaigns and education, promoting personal protective measures such as mask-wearing and regular hand hygiene. The implementation of these measures played a pivotal role in curbing the spread of the pandemic and safeguarding public health (10,11).

Beyond these fundamental steps, the Macao SAR government demonstrated flexibility in its approach to preventive and control measures. They enhanced rapid testing and contact tracing to promptly identify and isolate cases. Collaborative efforts with healthcare sectors, community leaders, and the public were strengthened to ensure a coordinated and effective response. Furthermore, the Macao SAR government swiftly adjusted its economic policies during the pandemic to alleviate the impact on affected businesses and individuals. By enacting a series of emergency policies and measures, the government provided financial support to affected industries while implementing measures to safeguard the basic needs of the public. This comprehensive policy response reflects the government's holistic approach to addressing the crisis, aiming to protect public health while maintaining overall societal stability.

Highlight box

Key findings

- During coronavirus disease 2019 (COVID-19) in the Macao Special Administrative Region (SAR), the government formulated and implemented a series of prevention and control policies. The government's decision-making process, policy formulation, and implementation strategies during the epidemic also need to take into account economic and social considerations.

What is known and what is new?

- The Macao SAR formulated many prevention and control policies during the period of COVID-19.
- During the COVID-19 period, the Macao SAR government's prevention and control strategies were carefully drafted to strike a balance between mitigating the epidemic and facilitating economic recovery. Additionally, the government acknowledged and addressed the public's mental health and educational needs. The government has made many efforts to meet the needs of public health and economic stability during the epidemic.

What is the implication, and what should change now?

- Conducting a thorough analysis of the vertical assessment of policy implementation and meticulously evaluating its enduring effects on public health, economic revival, and social resilience yields invaluable experiences and insights for managing future crises management.

The implementation of community lockdowns and vaccination strategies, while crucial in controlling the spread of COVID-19, carries potential negative impacts. Community lockdowns, for instance, can lead to feelings of isolation and helplessness among individuals, potentially resulting in mental health challenges. Furthermore, such lockdowns may disrupt the free flow of information, inadvertently fostering the spread of rumors and causing public panic. Another consequence of lockdowns is the potential erosion of public trust and support for health measures, which can diminish the effectiveness of epidemic prevention efforts. It is essential, therefore, to ensure clear and transparent communication to maintain public confidence and cooperation. Vaccination, a key strategy in controlling COVID-19, can elicit varying responses among different demographic groups. This variability necessitates the development of personalized vaccination strategies, tailored to the specific characteristics and needs of diverse populations. Such an approach can enhance vaccination coverage and effectiveness. To optimize the management of the pandemic, it is crucial to integrate socio-demographic and health factors in formulating targeted community lockdown strategies, vaccination plans, and other public health measures. This comprehensive and personalized approach aims to not only curb the virus's spread but also minimize the negative impacts on mental health, information dissemination, and public trust (12,13).

The political system of the Macao SAR is based on the principles of executive leadership, legislative constraints, and judicial independence. According to the "Macao Basic Law", the administration and legislation restrict and cooperate with each other. The government must be responsible to the Legislative Council and accept its supervision. The actions of the chief executive and the government are also subject to legal review by the courts. In order to prevent the local powers of the Macao SAR from being too powerful, the central government has retained some supervisory powers over its local autonomy. The Legislative Council has the authority to review and pass the government's financial budget, raise questions on government work, and exercise supervision over the government. Members of the Legislative Council are elected by direct election, indirect election and appointment. They represent the interests and demands of residents and work with the government to promote the formulation of laws and policies. During the COVID-19 epidemic, the Macao Legislative Council actively supervises the government's epidemic prevention measures, reviews and amends relevant

laws and policies, encourages senior government officials to explain the implementation and effects of measures, ensures transparency in decision-making, and requires the government to disclose epidemic information and prevention and control measures. Effectiveness, supervise the allocation of resources, and provide legal proceedings and remedies to protect public rights and interests. From the transparency and dynamics of the epidemic, to the discussion and interpretation of COVID-19 policies, to the effectiveness and social rationality of prevention and control measures, as well as the allocation of medical resources and service quality, the Legislative Council has played an important supervisory role in all aspects. This kind of supervision not only helps protect public health and safety, but also ensures the scientific nature, fairness and transparency of government prevention and control measures, further safeguarding social stability. Through these efforts, the Legislative Council's sense of responsibility and responsibility in responding to the epidemic has been demonstrated, and the public's understanding and trust in the work of the Legislative Council have also been enhanced.

Rationale and knowledge gap

Macao SAR's COVID-19 prevention and control policies extended beyond the immediate concerns of the epidemic to encompass broader economic and societal impacts. In the two phases of the COVID-19 pandemic, a meticulous analysis of keyword vectors reveals significant shifts in the response strategies of the Macao SAR government.

Objective

This study aims to comprehensively analyze the evolution of COVID-19 prevention and control policies in the Macao SAR. Through the analysis of speeches and inquiries from Macao SAR Legislative Assembly members, along with relevant news reports, we seek an in-depth understanding of the government's decision-making processes, policy formulation, and implementation strategies during the pandemic.

Methods

Datasets

The data collection for this study comprises two primary

components. Initially, information was sourced from the website of the Legislative Assembly of the Macao SAR of the People's Republic of China (14-16). The search focused on legislators' speeches and inquiries addressing COVID-19 prevention and control policies, specifically employing the search criteria with the phrase "supervision of epidemic prevention policies". The data extraction covered the period from January 20, 2020 to May 23, 2023, and involved searching for titles and content of speeches and inquiries using the "full-text search" and "include synonyms" options.

Subsequently, data was retrieved from the WiseSearch website. The search criteria incorporated keywords such as "Legislative Assembly", "legislator", and "epidemic prevention", along with terms like "COVID-19", "pandemic", or "nucleic acid". The search encompassed both title and content, and media categories included newspapers and websites, comprehensive news media types, and media regions related to Macao. The timeframe aligned with that of the Legislative Assembly website, spanning from January 20, 2020 to May 23, 2023.

Within the Legislative Assembly website, a total of 4,456 topics related to COVID-19 prevention and control were identified, featuring the keyword "supervision of epidemic prevention policies". These topics collectively comprised approximately 5.7 million words. Notably, the annual discussion of topics related to COVID-19 prevention and control in the Legislative Assembly of Macao SAR exhibited a steady increase. In 2020, 60 topics were discussed, a number that surged to 252 in 2021 as the pandemic evolved. By 2022, this figure further had escalated to 707 topics. In 2023, the number of topics related to COVID-19 prevention and control reached an unprecedented 1,435. In summary, over the past few years, a total of 4,456 topics related to COVID-19 prevention and control have been deliberated in the Legislative Assembly of the Macao SAR.

Word vectorization

Word vectorization plays a crucial role in natural language processing, enabling the use of mathematical methods to process and comprehend textual data. Given a vocabulary set (V) containing (N) words (w_1, w_2, \dots, w_n), word vectorization refers to mapping each vocabulary term (w_i) to a real-number vector space (R^d), denoted as ($v_i \in R^d$), where (d) represents the dimensionality of the word vector. This mapping captures semantic and syntactic relationships between words, allowing for efficient representation

and computation of vocabulary terms in a vector space. Word vectors primarily include methods such as One-hot representation and Distributed representation (17).

One-hot representation

One-hot representation is a method that represents each word in a vocabulary as a high-dimensional vector, where only one dimension has a value of 1 to indicate the presence of the word, while all other dimensions are 0. Assuming we have a vocabulary set ($V = v_1, v_2, \dots, v_N$) containing (N) different words, the One-hot representation for each word (w_i) is a (N)-dimensional vector (X_{w_i}), where only one dimension has a value of 1, and all other dimensions are 0. Specifically, for a word (w_i), its One-hot representation is:

$$X_{w_i} = [0, 0, \dots, 1, \dots, 0] \quad [1]$$

One-hot representation maps each word to a high-dimensional vector space equal to the size of the vocabulary set ((N)). Despite its simplicity, One-hot representation presents challenges in computation and storage due to its high-dimensional sparsity, making it difficult to capture semantic correlations between words.

Distributed representation

In contrast, Distributed representation employs a dense vector representation by mapping each word to a low-dimensional real-number vector space (R^d), where (d) is the dimensionality of the word vector. In this representation, the relative positions of word vectors for different terms in the vector space reflect their semantic and syntactic relationships. This approach more effectively discerns semantic similarity and syntactic structures among vocabulary terms, providing significant applications in natural language processing tasks.

Suppose we have a vocabulary set (V) containing (N) words, and we want to map each word (w_i) to a (d)-dimensional real-number vector space. Using a word embedding matrix ($E \in R^{N \times d}$) to represent all word distributed representations, where the (i)-th row represents the vector representation of the (i)-th word (w_i), i.e., ($E[i,:]$) represents the vector representation of the (i)-th word. For each word (w_i) in the vocabulary set (v), its word vector can be represented as ($v_i = E[i,:] \in R^d$).

To learn an appropriate word embedding matrix (E), a corpus ($D = X^{(1)}, X^{(2)}, \dots, X^{(M)}$) comprising (M) training samples, where each training sample ($x^{(m)}$) is composed of word sequences, is utilized. The objective is to adjust the embedding matrix (E) by minimizing a loss function to ensure that words with semantic similarities are closer in distance in the vector space, thereby better capturing semantic similarities and syntactic structures among vocabulary terms.

Common loss functions include softmax loss and negative sampling loss, both based on the similarity or relative position of word vectors.

Softmax loss is commonly used for training word vector models, assuming a context-target word pair ((c, o)), where (c) represents the context word vector and (o) represents the target word (or center word) vector. The objective is to minimize the softmax loss function to adjust the word embedding matrix (E), ensuring that the semantic similarity between context and target words is captured in the vector space.

Let (v_c) and (v_o) represent the word vectors for the context and target words, respectively. Their dot product reflects their similarity in vector space. The softmax loss function is formalized as:

$$L(\theta) = -\log\left(\frac{\exp(v_c \cdot v_o)}{\sum \exp(v_c \cdot v_i)}\right) \quad [2]$$

where ($v_c \cdot v_i$) represents the dot product of the context word vector (v_c) with the vectors of all words in the vocabulary, ($\exp(x)$) represents (e) raised to the power of (x). Minimizing the softmax loss function enables learning an appropriate word embedding matrix (E) such that the dot product values of semantically similar words are closer in the vector space.

Negative sampling loss is another commonly used form of loss function that reduces computational complexity by introducing negative samples for each training sample. Its formal description is as follows:

$$L(\theta) = -\log(\sigma(u_o^T v_c)) - \sum_{i=1}^k \log(\sigma(-u_i^T v_c)) \quad [3]$$

where, ($\sigma(x)$) represents the Sigmoid function, (u_o)

denotes the word vector for the target word (o), (v_c) represents the word vector for the context word (c), (u_i) represents the word vector for the negatively sampled word (i), and (k) is the number of negative samples. In the negative sampling loss function, the first term ensures that the dot product value for the positive sample approaches 1, resulting in a smaller loss. The second term ensures that the dot product values for negative samples approach 0, contributing to a smaller loss. By minimizing the negative sampling loss function, the word embedding matrix (E) is adjusted to increase the dot product values for positive samples and decrease those for negative samples, thereby more accurately capturing semantic similarities and syntactic structures among vocabulary terms. By optimizing the loss function, an appropriate word embedding matrix (E) can be learned, enhancing the representation of words in the vector space to better reflect their semantic and syntactic relationships.

Word vector models

Word vector models play a crucial role in natural language processing, providing a means to map words into a continuous vector space, thereby enabling the application of mathematical techniques for text data analysis and comprehension. Among these models, Word2Vec is commonly employed, leveraging neural network methodologies to learn word vectors from extensive corpora (18).

Assume a vocabulary list V containing N words w_1, w_2, \dots, w_n , where each word w_i is represented as a vector $v_i \in \mathbb{R}^d$ in the d -dimensional vector space. The objective is to learn appropriate word vectors by minimizing the following objective function:

$$J(\theta) = \frac{1}{T} \sum_{t=1}^T \sum_{-m \leq j \leq m, j \neq 0} \log P(w_{t+j} | w_t) \quad [4]$$

where T denotes the total number of training samples, m signifies the size of the context window, i.e., the number of words surrounding the target word considered. $P(w_{t+j} | w_t)$ represents the probability of the occurrence of the context word w_{t+j} given the target word w_t . Word2Vec model computes $P(w_{t+j} | w_t)$ using two distinct approaches:

(I) Skip-Gram model: given a target word w_t , it aims

to predict its surrounding context words w_{t+j} . The Skip-Gram model calculates $P(w_{t+j} | w_t)$ by maximizing the following conditional probability:

$$P(w_{t+j} | w_t) = \frac{\exp(v_{t+j}^T v_{w_t})}{\sum_{i=1}^N \exp(v_i^T v_{w_t})} \quad [5]$$

where v_i denotes the word vector of word w_i .

(II) CBOW model: given a target word w_t , it seeks to predict it based on its context words w_{t+j} . The CBOW model computes $P(w_t | w_{t+j})$ by maximizing the following conditional probability:

$$P(w_t | w_{t+j}) = \frac{\exp(v_{w_t}^T \sum_{-m \leq i \leq m, i \neq 0} v_{w_{t+i}})}{\sum_{i=1}^N \exp(v_i^T \sum_{-m \leq i \leq m, i \neq 0} v_{w_{t+i}})} \quad [6]$$

where v_i symbolizes the word vector of word w_i .

Through the optimization of the objective function $J(\theta)$, we can fine-tune word vectors and model parameters via backpropagation. This process culminates in the acquisition of optimal word vector representations. At its core, Word2Vec is a renowned word vector model that gleans insights from extensive corpora to learn word vectors. This optimization of the objective function allows for the adjustment of word vectors, thereby bringing semantically similar words closer in the vector space. The resulting word vectors are of considerable value in the realm of natural language processing tasks.

Performance evaluation metrics

In word vector models, various methods can be utilized to calculate the similarity between word vectors, among which cosine similarity is a common measure for assessing the degree of similarity between two-word vectors, gauging their directional similarity by computing the cosine distance (19).

Given two-word vectors x and y , the cosine similarity is calculated as follows:

$$\cos(x, y) = \frac{x \cdot y}{|x| \cdot |y|} \quad [7]$$

where x and y respectively represent the word vectors of two words, \cdot denotes the dot product, and $|x|$ and $|y|$ signify the Euclidean lengths of the vectors.

Cosine similarity values range from -1 to 1 . When two-word vectors align in the same direction, the cosine similarity is 1 ; when they are diametrically opposed, the cosine similarity is -1 ; and when they are nearly orthogonal, the cosine similarity approximates 0 . The calculation of cosine similarity has the following characteristics:

- (I) Cosine similarity remains unaffected by the magnitude of word vectors and is solely influenced by their directional orientation. This means that word vectors can achieve high cosine similarity if their directional orientations are similar, regardless of differences in their lengths.
- (II) Cosine similarity is a symmetric similarity calculation method, i.e., $\cos(x, y) = \cos(y, x)$. This implies that the calculation result of the similarity between two-word vectors is not affected by their order.

In word vector models, cosine similarity can be employed to compare the degree of similarity between two words. Generally, the closer the cosine similarity of two words is to 1 , the more similar their semantic and syntactic relationships are; the closer the cosine similarity is to -1 , the more opposite their semantic and syntactic relationships are; when the cosine similarity is close to 0 , it indicates a weak semantic and syntactic relationship.

Experiment setting

In this study, we employed four commonly used Chinese segmentation tools (jieba, SnowNLP, THULAC, NLPIR) and compared their differences in text segmentation performance (20).

Firstly, we wrote the code for the four segmentation tools using Python 3.5, and used text as input to perform segmentation processing through these tools. Subsequently, we carried out a comparative analysis of the segmentation results to determine which tool could provide more accurate and efficient segmentation results.

In terms of segmentation tool selection, we found that the THULAC tool performed excellently with strong capabilities, high accuracy, and speed, and it also supports custom dictionaries. Therefore, we chose THULAC as the segmentation method for this study and further researched the impact of custom dictionaries on segmentation accuracy.

Apart from the choice of segmentation tools, this study also adopted Word2Vec as the word vector model. Based on this, we used the Word2Vec tool in the Gensim topic model

package for training the word vector model and calculating similarities. For the CBOW model and Skip-Gram model, we chose Word2Vec in Gensim for experimentation and comparative analysis.

In Table S1, lines 1 to 2 are the codes for jieba word segmentation, and the results are saved in the words variable. Lines 3 to 6 of code are NLPIR word segmentation codes, and the results are stored in the segments variable. The 7th to 8th lines of code are SnowNLP word segmentation codes, and the results are stored in the s variable. Lines 9 to 11 of code are THULAC word segmentation codes, and the results are saved in the thulac_word_seg.txt file. Run the obtained word segmentation results for subsequent operations.

In terms of word segmentation tool selection, we found that the THULAC tool performed well, with strong capabilities, high accuracy, fast speed, and it also supports custom dictionaries. Therefore, we choose THULAC as the segmentation method for this study and further study the impact of custom dictionaries on segmentation accuracy.

In addition to the selection of word segmentation tools, this study also uses Word2Vec as the word vector model. Based on this, we use the Word2Vec tool in the Gensim topic model package to train the word vector model and calculate the similarity. For the CBOW model and Skip-Gram model, we choose Word2Vec in Gensim for experiments and comparative analysis.

In order for the program to use the Gensim tool, the first line of code imports models and similarities from the Gensim library. The news text data will be read from the word_seg.txt file into the variable texts through the 2nd to 3rd lines of Table S2, and will be steadily divided into corpus arrays through the 4th line of code. The next 5 to 17 lines of code are to count the word frequency of the corpus, and the 18 to 20 lines are to build the Word2Vec model of the corpus and calculate the similarity. The 5th to 10th lines of code are to count the frequency of the news corpus and write it into the dictionary variable count_dict. Then, in the 11th line of code, the dictionary variable count_dict is sorted from high to low by frequency. Finally, the 12th to 17th lines of code are to sort the sequential word dictionary is saved to the result.txt file. Line 18 is to build a Word2Vec model for the corpus. Lines 19 to 20 are based on the model to calculate and output the similarity of the two-word vectors “COVID-19” and “Epidemic control measure”.

During the experiment, we compiled a custom dictionary, marked and extracted fixed phrases through

repeated segmentation experiments and reading related literature, to construct a terminology library suitable for this study. Afterwards, we used the Gensim toolkit to count word frequencies in the text, built a Word2Vec model, and calculated the similarity between word vectors to obtain experimental results.

Results

Keyword frequency analysis of COVID-19 prevention and control policies in Macao SAR

Table 1 presents the results of keyword frequency analysis of COVID-19 prevention and control policies in Macao SAR. Through the examination of speeches and inquiries from Macao SAR legislative council members, the following analyses were derived:

In the realm of pandemic control measures, the keyword “Vaccination” emerged as the most frequently mentioned in the analysis (occurring 338 times) and “COVID-19” (occurring 297 times). This underscores the significance of vaccine administration and the attention given to COVID-19 during the pandemic control process. Additionally, other notable keywords such as “Tourism” (294 occurrences), “Anti-epidemic” (253 occurrences), and “Infection” (161 occurrences) also received considerable attention.

In terms of government response, “Government” was the most frequently used keyword (occurring 1,082 times), indicating the pivotal role played by the government in pandemic response. Other keywords such as “Employment” (324 occurrences), “Policy” (243 occurrences), and “Support” (198 occurrences) underscored the government’s focus on employment, policy formulation, and support measures. The specific keyword “Port clearance” (116 occurrences) may indicate policies related to port clearance.

In terms of economic impact, keywords such as “Economy” (855 occurrences) and “Unemployment” (217 occurrences) stood out, highlighting the significant impact of COVID-19 on Macao SAR’s economy and the exacerbation of unemployment issues. Furthermore, “On duty” (78 occurrences) and “Subsidy” (30 occurrences) reflected attention to on-duty personnel and economic subsidies.

In the healthcare system domain, “Healthcare” (114 occurrences) and “Community” (137 occurrences) were the most frequently used keywords, emphasizing the focus on healthcare and community-level preventive measures.

Table 1 Analysis of key COVID-19 terms in the context of the evolution of Macao SAR's pandemic control policies

Field	Vocabulary	Frequency
Epidemic control measures	Vaccination	338
	COVID-19	297
	Tourism	294
	Anti-epidemic	253
	Infection	161
	Nucleic acid	160
	Quarantine	134
	Pneumonia	125
	Case	101
	Health code	94
	Diagnosis	83
Government response	Government	1,082
	Employment	324
	Policy	243
	Support	198
Economic impact	Port clearance	116
	Unemployment	217
	Economy	855
	On duty	78
Healthcare system	Subsidy	30
	Expenditure	23
	Community	137
	Healthcare	114
Community services	Critical condition	27
	Medical observation	18
	Student	129
	Psychological	49
	Social service	7

COVID-19, coronavirus disease 2019; Macao SAR, Macao Special Administrative Region.

Meanwhile, “Expenditure” (23 occurrences) and “Critical condition” (27 occurrences) were related to resource allocation and severe cases.

Frequent keywords in community services included “Student” (129 occurrences) and “Psychological” (49 occurrences), highlighting attention to community services

for students and psychological support.

The analysis reveals the Macao SAR government's high attention to vaccination, pandemic control measures, and economic impact throughout the evolution of COVID-19 prevention and control policies. The government responded by formulating policies, providing support, and addressing port clearance. Additionally, Macao SAR prioritized healthcare system development, strengthened community services, especially support for students and mental health.

Word vector and keyword similarity analysis results

The examination of word vector similarity in relation to the progression of COVID-19 prevention and control policies in Macao SAR revealed notable trends and variations. Particularly, when considering the roles of the legislative council and the Statistics and Census Service (DSEC), *Table 2* illustrates the vector space similarity values between different departments and keywords.

Legislative Council's role and influence

High policy relevance: the legislative council exhibited high similarity with keywords such as “Measures” (0.84368706), “Government” (0.8400955), and “Society” (0.84198743), indicating its central role in policy formulation, government coordination, and societal management during the pandemic. This strong correlation signifies the legislative council's leadership and decision-making role in shaping and executing pandemic control strategies.

Central position in pandemic control

The legislative council demonstrated high similarity with “Epidemic” (0.8382773), “Prevention” (0.8371937), and “Virus” (0.8361382), emphasizing its crucial position in pandemic control and responding to virus spread. This suggests that the legislative council plays a key role not only in policy formulation but also in the implementation and management of specific epidemic control measures.

Comprehensive consideration of vaccination policies

While the similarity with “Vaccination” (0.8107879) and “Vaccine” (0.7858573) was relatively lower, it does not imply a lack of influence. This data may indicate that the legislative council took a comprehensive approach in considering vaccination as part of broader pandemic control strategies, with detailed vaccination policies potentially overseen by other specialized departments.

Table 2 Analysis of word vector and keyword similarity in the context of the evolution of Macao SAR's pandemic control policies

Government department	Vocabulary	Cosine similarity
Legislative Council	Epidemic	0.8382773
	Government	0.8400955
	Measures	0.84368706
	Residents	0.8345191
	Preventive Measures	0.8371937
	Vaccination	0.8107879
	Vaccine	0.7858573
	Safeguard	0.83690643
	Society	0.84198743
	Virus	0.8361382
Statistics and Census Bureau	Epidemic	0.22247018
	Government	0.21863604
	Measures	0.22811559
	Residents	0.20780647
	Preventive Measures	0.2152394
	Vaccination	0.19039541
	Vaccine	0.2173695
	Safeguard	0.22138484
	Society	0.22424719
	Virus	0.21474826

Macao SAR, Macao Special Administrative Region.

Role of the DSEC

Data analysis and support: the Statistics and Census Service showed generally low similarity with all keywords, such as “Epidemic” (0.22247018), “Government” (0.21863604), and “Measures” (0.22811559). This suggests a limited role in direct policy formulation or execution but a focus on data collection and analysis, providing essential support for pandemic control strategies.

Impact in specific areas

Higher similarity values, such as with “Measures” (0.22811559) and “Virus” (0.21474826), imply that the Statistics and Census Service played a role in providing data support for epidemic control measures and statistical analysis of virus spread. However, lower similarity with “Vaccination” (0.19039541) and “Vaccine” (0.2173695) suggests limited involvement in vaccination strategy and

execution, focusing more on providing data support for these areas.

Collaborative efforts between departments

The legislative council took a leading role in drafting and coordinating pandemic control policies, evident from its high similarity with keywords, emphasizing its extensive involvement and influence. In contrast, the Statistics and Census Service focused primarily on providing data support and analysis, with generally lower similarity to keywords. These findings highlight the legislative council's importance in comprehensive policy formulation and the Statistics and Census Service's role as a crucial provider of information and data support. These discoveries underscore effective cooperation and functional complementarity among government departments, ensuring the comprehensiveness and effectiveness of policy implementation.

Analysis of two waves of the pandemic

In this study, a detailed analysis of the evolution of COVID-19 prevention and control policies in Macao SAR during two waves of the pandemic (2021.06–2021.10 and 2022.06–2022.07) was conducted as shown in *Figure 1*. Utilizing radar charts as a form of data visualization, a comparative analysis of keyword frequencies during the first and second waves was performed. The radar chart presents keyword frequencies in a circular manner, with each keyword positioned at a specific angle, and the distance from the center indicating its frequency. The first wave data is represented with a green border and gray fill, whereas the data from the second wave is depicted. This visual comparison vividly illustrates changes in keyword frequencies during the two time periods.

The frequency changes of key terms provide insights into the evolving strategies employed by the Macao SAR during the two waves of the COVID-19 pandemic. Notably, the term “Epidemic”, which accounted for a substantial 0.196 during the first wave, experienced a decline to 0.105 in the second wave. This shift indicates a gradual transition in public and decision-makers' focus from the epidemic itself towards specific preventive measures, reflecting a deepening understanding of the pandemic.

The term “Government” maintained a consistently high frequency in both waves, registering at 0.155 and 0.094, underscoring the crucial and enduring role played by the government in responding to the pandemic. The term “Epidemic Prevention” held significant proportions in both

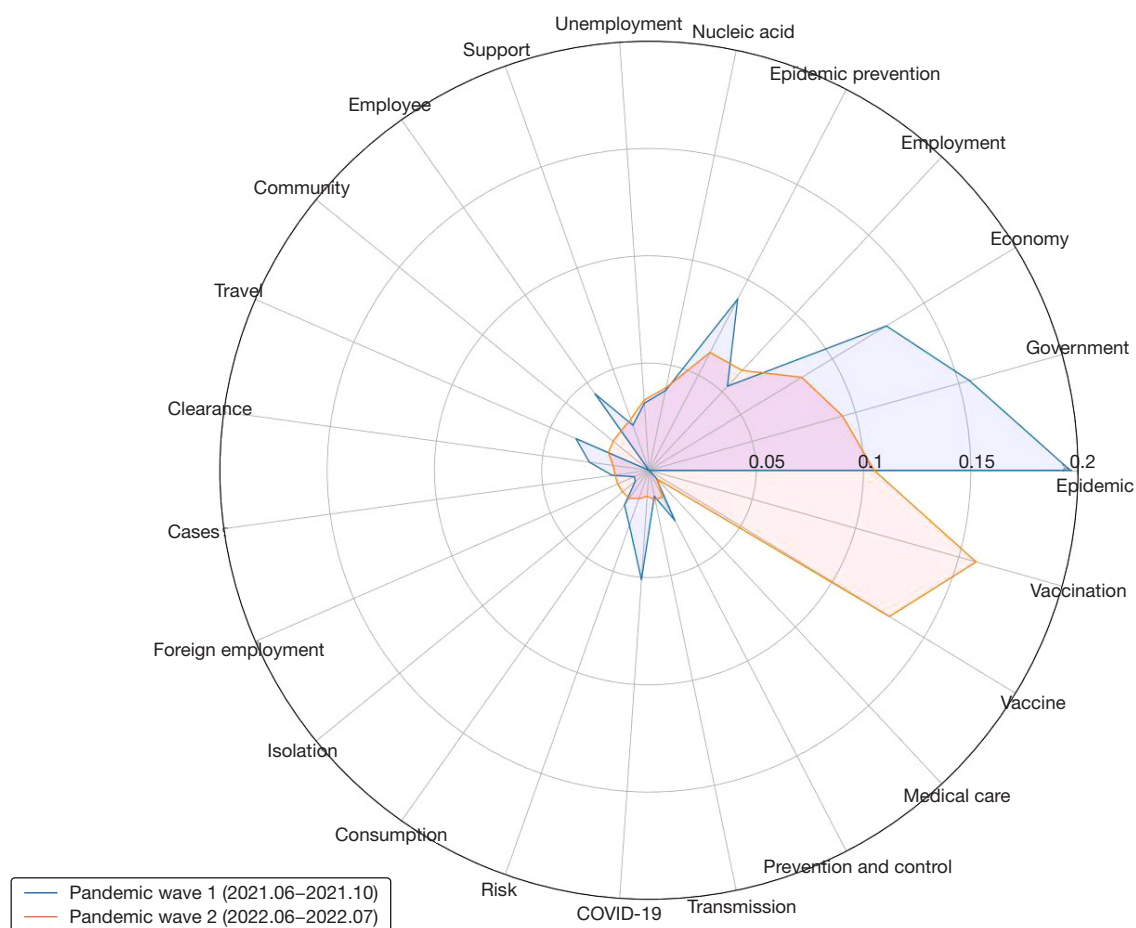


Figure 1 Radar analysis of policy evolution in Macao SAR during two pandemic waves.

waves, at 0.090 and 0.062, emphasizing the continuous commitment of the Macao SAR to epidemic prevention and control, mainly by issuing announcements and guidelines on the epidemic, providing protective measures and relevant information to the public.

Keywords such as “Travel” and “Quarantine” exhibited a relative decrease in frequency during the second wave, at 0.021 and 0.016, respectively, reflecting a diminished emphasis on travel-related and quarantine policies during this period.

Conversely, “Vaccine” and “Vaccination” emerge as focal points in the second wave, with proportions of 0.131 and 0.158, respectively. This signals heightened attention from Macao SAR toward specific prevention measures such as vaccination, to improve the immune level of the population and reduce the risk of epidemic transmission.

The keywords “Economy” and “Employment” experienced an increase in frequency in the second

wave, at 0.083 and 0.064, respectively. This emphasizes the commitment of the SAR to balance economic and employment considerations during the epidemic, and reflects that the effectiveness and affinity of the COVID-19 prevention and control policies are two important dimensions for evaluating its comprehensive quality. Ensuring that policies are based on scientific evidence and undergo professional evaluation can help improve the effectiveness of measures. At the same time, comprehensive strategies and cross departmental and cross regional coordination and cooperation are also key to achieving effective prevention and control. In terms of people-friendliness, policies should focus on social equality and ensure that the interests of all groups are balanced, especially the needs of vulnerable groups. Information transparency and social support are also the cornerstone of people-oriented policies, which can enhance public understanding and cooperation, and provide necessary

assistance and support to people. Most importantly, policy formulation should fully listen to the public's voice and reflect democratic principles. On this basis, COVID-19's prevention and control policy can truly achieve the goal of protecting public health, respecting human rights and addressing global challenges.

The fluctuation in keyword frequencies unveils the strategic evolution of Macao SAR's response to the two waves of COVID-19. Moving from initial emergency responses to more mature and comprehensive measures, the strategy encompasses considerations for various aspects of the economy and society. This dynamic evolution of the Zero-COVID strategy reflects Macao SAR's continuous optimization and adjustment in the face of persistent challenges.

Discussion

Our study employed a combination of various Chinese word segmentation tools and the Word2Vec word embedding model, offering a novel perspective for a profound understanding of policy texts. The integration of THULAC word segmentation and Word2Vec allowed not only the precise capture of keyword usage frequency but also revealed intricate relationships among different keywords. This approach demonstrated excellent performance in handling large-scale textual data, providing an efficient tool for policy analysis.

Using the aforementioned methodology, our study delved into a detailed analysis of the evolution of COVID-19 prevention and control policies in Macao SAR. By comprehensively examining statements from legislators and relevant policy documents, we identified rapid and effective measures taken by the Macao government in response to the pandemic. Key areas of focus included vaccination strategies, adjustments to the tourism industry, and the implementation of social distancing measures. Notably, strict border controls and health monitoring measures adopted by the Macao government in the early stages of the pandemic contributed to controlling the rapid spread of the virus, albeit imposing economic pressures, particularly on the tourism-dependent local economy.

The frequency analysis revealed the government's heightened attention to keywords such as "Vaccination" and "COVID-19", emphasizing the pivotal importance of vaccination strategies within the overall prevention and control measures. Concurrently, the increased discussion frequency related to economic and employment concerns

reflected the government's growing consideration of the economic impact within its pandemic response strategy.

Furthermore, the analysis of word vectors and keyword similarity underscored the pivotal role of the Legislative Assembly in policy formulation. The high similarity between the Legislative Assembly and keywords such as "Government", "Measures", and "Society" highlighted its active involvement and significant influence in framing pandemic control policies, government coordination, and social management. This involvement extended beyond policy formulation to the execution and management of specific epidemic prevention measures.

Considerations for economic and social aspects

Macao SAR's COVID-19 prevention and control policies extended beyond the immediate concerns of the epidemic to encompass broader economic and societal impacts. The frequencies of keywords such as "Unemployment" and "Economy" indicated the significant economic ramifications of the pandemic on Macao SAR, prompting governmental efforts to alleviate these effects. The equilibrium achieved in the policies, simultaneously addressing epidemic prevention and economic recovery, demonstrated a thorough contemplation of both immediate and extended impacts. Simultaneously, the frequencies of keywords related to community services, such as "Student" and "Psychological", indicated the government's recognition of the psychological well-being and educational needs of the population. This holistic approach to addressing various societal needs reflected the government's comprehensive and humane crisis management.

Policy changes during two pandemic waves

In the two phases of the COVID-19 pandemic, a meticulous analysis of keyword vectors reveals significant shifts in the response strategies of the Macao SAR government. During the second wave, the government not only strengthened epidemic prevention and control measures but also notably heightened its focus on economic recovery and employment issues. This underscores its determination to strike a balance between public health safety and socio-economic recovery. Particularly noteworthy is the government's profound emphasis on "Vaccine" and "Vaccination", with frequencies accounting for 0.131 and 0.158, respectively. This signifies a deep understanding of specific preventive measures, showcasing flexibility and diversified thinking in

the face of sustained pandemic pressures. Simultaneously, the increased frequencies of the keywords “Economy” and “Employment” affirm the government’s efforts to navigate the needs for both public health and economic stability during the pandemic. Therefore, the strategic adjustments by the Macao SAR government in response to the pandemic not only reflect its adept handling of emergencies but also demonstrate a wise and decisive response to the continuously evolving situation. This ongoing optimization of the Zero-COVID strategy displays the government’s ability to comprehensively consider demands across various dimensions, including health, economy, and society.

Conclusions

In conclusion, our study offers profound insights into the COVID-19 pandemic response in Macao SAR, revealing intricate details of the government’s strategies. Furthermore, it underlines the efficacy of data-driven methodologies in shaping public policy analyses. Importantly, our findings furnish valuable experiences and insights essential for future crisis management. Moving forward, prospective investigations may delve into a longitudinal assessment of policy implementations, scrutinizing their enduring impacts on public health, economic resurgence, and societal resilience. Moreover, investigating the dynamic realm of emerging technologies, including artificial intelligence and big data, and their potential to enhance policy responses and crisis management, presents a valuable direction for future research in this field.

Acknowledgments

Funding: This work was supported by the Science and Technology Development Fund of Macau SAR (005/2022/ALC), the Self-Supporting Program of Guangzhou Laboratory (SRPG22-007), and the National Basic Research Priorities Program of China (grant No. 2023YFC3041600).

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editors (Jing Cheng, Tao Xu, Zifeng Yang, and Wenda Guan) for the series “Current Status of Diagnosis and Forecast of COVID-19” published in *Journal of Thoracic Disease*. The article has undergone external peer review.

Data Sharing Statement: Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-23-1818/dss>

<https://jtd.amegroups.com/article/view/10.21037/jtd-23-1818/dss>

Peer Review File: Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-23-1818/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-23-1818/coif>). The series “Current Status of Diagnosis and Forecast of COVID-19” was commissioned by the editorial office without any funding or sponsorship. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Hu B, Guo H, Zhou P, et al. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 2021;19:141-54. Erratum in: *Nat Rev Microbiol* 2022;20:315.
- Li J, Lai S, Gao GF, et al. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* 2021;600:408-18.
- Burki T. Dynamic zero COVID policy in the fight against COVID. *Lancet Respir Med* 2022;10:e58-9.
- Zhao YJ, Xu JQ, Bai W, et al. COVID-19 prevention and control strategies: learning from the Macau model. *Int J Biol Sci* 2022;18:5317-28.
- Ieng SM, Cheong IH. An Overview of Epidemiology of COVID-19 in Macau S.A.R. *Front Public Health* 2020;8:550057.
- Chau KY, Luo JM, Duan X. A Qualitative Investigation of the Impact of COVID-19 on Macau’s Gaming Industry. *Journal of Quality Assurance in Hospitality & Tourism* 2022;23:639-49.

7. Ung COL, Hu Y, Hu H, et al. Investigating the intention to receive the COVID-19 vaccination in Macao: implications for vaccination strategies. *BMC Infect Dis* 2022;22:218.
8. Xinhua News. Macao starts vaccination of mainland-made COVID-19 vaccines, chief executive goes first. 2021. 13 January 2023. Available online: http://www.xinhuanet.com/english/2021-02/09/c_139732367.htm
9. Macao CDC. Special webpage against Epidemics-COVID-19 Vaccination. 2023. 13 Jan 2023. Available online: <https://www.ssm.gov.mo/apps1/PreventCOVID-19/ch.aspx#clg22916>
10. Liang J, Liu R, He W, et al. Infection rates of 70% of the population observed with-in 3 weeks after release of COVID-19 restrictions in Macao, China. *J Infect* 2023;86:402-4.
11. Zeng Z, Qu W, Liu R, et al. Real-time assessment of COVID-19 epidemic in Guangdong Province, China using mathematical models. *J Thorac Dis* 2023;15:1517-22.
12. Gosak M, Duh M, Markovič R, et al. Community lockdowns in social networks hardly mitigate epidemic spreading. *New J Phys* 2021;23:043039.
13. Markovič R, Šterk M, Marhl M, et al. Socio-demographic and health factors drive the epidemic progression and should guide vaccination strategies for best COVID-19 containment. *Results Phys* 2021;26:104433.
14. Macau Legislative Council written consultation. Available online: <https://www.al.gov.mo/zh/written-consultation>
15. Macau Legislative Council oral consultation. Available online: <https://www.al.gov.mo/zh/oral-consultation/search>
16. Macau Legislative Council Speaking before agenda. Available online: <https://www.al.gov.mo/zh/speaking>
17. Church KW. Word2Vec. *Natural Language Engineering* 2017;23:155-62.
18. Ma L, Zhang Y. Using Word2Vec to process big text data. In: 2015 IEEE International Conference on Big Data (Big Data); IEEE 2015:2895-7.
19. Rong X. word2vec parameter learning explained. 2014 arXiv preprint arXiv:1411.2738.
20. Goldberg Y, Levy O. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. 2014. arXiv preprint arXiv:1402.3722.

Cite this article as: Wong K, Wang B, Hsieh K, Hon C, Zeng Z, Ngai L, Liu Z, Wang Y, Hon S, Lao H, Lu G. Text analysis of Macao's COVID-19 prevention and control policies: discussion on strategy evolution and public health capabilities. *J Thorac Dis* 2024;16(1):632-644. doi: 10.21037/jtd-23-1818

Table S1 Calls four different word segmentation methods

```

1:         import jieba
2:         words = jieba.lcut(text)
3:         import pynlpir
4:         pynlpir.open()
5:         segments = pynlpir.segment(text,pos_tagging=False,pos_names='all', pos_english=False)
6:         pynlpir.close()
7:         from snownlp import SnowNLP
8:         s = SnowNLP(text).words
9:         import thulac
10:        thulac = thulac.thulac(user_dict="dict.txt", seg_only=True)
11:        thulac.cut_f("input.txt", "tmp\\thulac_word_seg.txt")

```

Table S2 Gensim performs Word2Vec modeling and similarity calculation

```

1:         from Gensim import models, similarities

```

```

2:         file_object = open("tmp\\word_seg.txt", encoding='UTF-8')
3:         sentence = file_object.read()
4:         texts = [[word for word in sentence.split()]]
5:         count_dict=[]
6:         for words in texts[0]:
7:             if(len(words) >= 2):
8:                 if(words not in count_dict):
9:                     count_dict[words] = 0
10:                    count_dict[words] += 1
11:                    count_rs = sorted(count_dict.items(), key=lambda x:x[1])
12:                    save_count =''
13:                    size = len(count_rs)
14:                    for i in range(size):
15:                        save_count += (count_rs[size-i-1][0] + " : " + str(count_rs[size-i-1][1]) + "\n")
16:                    save_object = open("result.txt", 'w', encoding='UTF-8')
17:                    save_object.write(save_count)
18:                    model = models.Word2Vec(texts, size=100, window=5, min_count=5)
19:                    similarity = model.similarity("COVID-19", "Epidemic control measure")
20:                    print("similarity = %.4f" % similarity)

```
