RESEARCH

The symbiotic effect of online searches and vaccine administration—a nonlinear correlation analysis of baidu index and vaccine administration data

Yixin Liu[1], Lingshi Ran[1], Yang Wang[1] and Yixue Xia[1*]

Abstract

This study primarily addresses the analytical problem of the mathematical mechanism underlying the associative impact between online searches and vaccine uptake, a relationship that has become increasingly relevant in the context of public health management. As internet search behaviors reflect public interest and sentiment, understanding their impact on vaccination trends is crucial for real-time health decision-making. A Logistic model is constructed to observe the fundamental evolutionary patterns between online searches and vaccine uptake. To explore their mutual influence, an impact function is defined, and the common structural factors with the highest fitness are determined through data fitting. Subsequently, a dynamic detection model of the associative impact between online data and societal objects, based on the mathematical mechanism, is established. Using this model, dynamic predictions are conducted to verify its predictive capability at certain stages. Through research, a symbiotic effect between online searches and vaccine uptake is identified, revealing a nonlinear correlation between the two. The model demonstrates the ability to predict vaccine uptake trends based on online search data, with certain prediction windows showing high accuracy. This research not only clarifies the mathematical mechanism underlying this relationship but also demonstrates the advantage of integrated analysis and prediction. It provides a new method for predicting online searches and vaccine uptake, offering theoretical and empirical support for public health and social science research.

Keywords COVID-19 vaccination, Web search, Mathematical mechanism, Symbiotic effect, Logistic model

Introduction

In today's societal systems, the entities that are interconnected can be roughly divided into two categories: network data and social objects. Network data refers to all digital information generated online that can be quantitatively analyzed. Social objects, on the other hand, constitute the tangible, foundational, and highly crystallized portion of social reality offline. They are the fundamental constituents of society, forming an integrated whole with specific quantities, qualities, modes of combination, and degrees of consolidation. Social objects possess tangible power to instigate change.

With the construction of an interactional world of virtuality and reality, the interaction between network data and social entities has become more complex. Many studies have explored the correlation between the two, such as predicting trends in social entities using network data, uncovering influencing factors, and measuring the

*Correspondence:
Yixue Xia
rcofnpog@126.com
[1] Research Center of Network Public Opinion Governance, China People's Police University, Langfang 065000, China

Liu *et al. BMC Public Health*   (2025) 25:975

Page 2 of 18

degree of influence. Among these, the analysis of the mathematical mechanism behind their correlation is still worth exploring. Clarifying the mathematical mechanisms, especially constructing continuous models evolving over time, will increase the interpretability of the correlation, endowing it with a philosophical foundation such as social evolution mechanisms, and thus achieving a combination of data-driven and theory-driven appro4.1aches. Particularly in enhancing interpretability, correlation detection based on mathematical mechanisms offers the advantage of integrated analysis and prediction.

In recent years, with the rapid development and widespread adoption of Internet technology, there has been a multitude of ways in which the public accesses information. Among these, internet search engines stand out as one of the primary channels through which the public obtains the latest news. They have increasingly played a vital role across various societal domains, particularly in the realm of public health. The outbreak of the COVID-19 pandemic at the end of 2019, as an unprecedented global health crisis, swiftly spread worldwide, resulting in heightened public health awareness and an increased demand for information. Concurrently, vaccination emerges as the most economically effective method for preventing, mitigating, and controlling infectious diseases [1], playing a pivotal role in disease control and prevention efforts. With the spread of the COVID-19 pandemic, there arises an urgent need for COVID-19 vaccination. Since the publication of the genetic sequence of the novel coronavirus in January 2020, by December 2020, multiple COVID-19 vaccines had been approved for emergency use globally. Critical issues such as misinformation, vaccine hesitancy, and accessibility have created significant barriers to achieving widespread vaccine uptake [2]. Misinformation has been a significant barrier to public health efforts [3], as incorrect or misleading information about vaccine safety and efficacy has influenced public perceptions and behavior. Additionally, vaccine hesitancy remains a critical challenge, particularly in regions where distrust in medical institutions is prevalent, or where there are logistical challenges in vaccine distribution [4]. Addressing these challenges is essential for improving global vaccination rates and controlling the spread of infectious diseases. However, a narrow public health perspective may not provide comprehensive solutions; instead, a broader approach that considers societal, behavioral, and informational factors is needed. Internet data, reflecting the complex nature of public attitudes, can offer valuable insights, as search engines have become a key source of information for the public regarding vaccination. During the vaccination process, the public's search for vaccine-related information may impact their attitudes and behaviors towards vaccination. Concerning COVID-19 vaccines specifically, their development, research, production, and application are all influenced by internet data, which in turn affects the evolution of internet data.

In this study, we quantify the ubiquitous internet data as online searches, specifically utilizing Baidu search index to represent the changes and trends in internet data. Simultaneously, we represent social entities with vaccination rates, specifically quantifying this as the number of COVID-19 vaccine doses administered in China. The aim of this research is to explore the underlying mechanisms and patterns of influence between internet searches and vaccine administration. This will enable a deeper understanding of the interplay between internet data and social entities, uncovering potential associative effects between internet searches and vaccination behaviors. By analyzing these associative effects, we aim to provide insights into future trends and predictions, thereby offering new theoretical and empirical support for the fields of public health and social sciences.

Related work
Online: internet searches
When people have specific needs, they engage in search behavior by entering relevant keywords into search engines to obtain information. These search behaviors are recorded by search engines and presented in the form of internet search data. Ginsberg et al. pioneered the use of internet search data for predicting epidemics, and their research demonstrated that models constructed using internet search data could forecast the outbreak of infectious diseases two weeks in advance [5]. Since then, internet search data have gradually become an important resource for scholars to study correlations and make predictions. Among them, Google Trends and Baidu Index are commonly used tools for internet search data, which have been widely applied in fields such as travel demand [6], disease monitoring [7], housing price changes [8], and policy implementation [9]. By analyzing trends and changes in internet search data using machine learning algorithms, researchers can effectively predict and evaluate offline entities. During the pandemic, studies have also used internet search data to predict and monitor the dosage of COVID-19 vaccinations. Compared to traditional survey and measurement methods, internet search data can provide more timely and comprehensive results, and are more correlated with the results obtained from traditional vaccination research methods [10]. Therefore, internet search have broad prospects for application in the field of vaccination.

Researchers employ various methods for analyzing internet search data. A review study by Nuti et al.

Liu *et al. BMC Public Health*   (2025) 25:975

Page 3 of 18

indicated that among 70 studies conducted between 2009 and 2013, 70% utilized time trend analysis (comparing across time periods), 11% conducted cross-sectional analysis (comparing across different locations within a single time period), and 19% of the studies employed both methods simultaneously [11]. These three approaches have been applied in studies related to vaccination. For instance, Pullan and Dey observed the popularity of search terms related to COVID-19 vaccines in Google Trends, suggesting that internet searches can help monitor public attitudes towards vaccines during rapidly changing global health crises such as the COVID-19 pandemic [12]. Awijen et al. collected samples from 194 countries worldwide, and their difference-in-differences investigation approach revealed that with the arrival of vaccines, trends in Google searches measuring fear and anxiety are increasing [13]. Moussa and Moussa found that utilizing Google Trends data, from 2004 to 2017, obtaining online vaccination information in sub-Saharan African countries influenced vaccination rates [14].

Researchers combine various methods to analyze internet data, including correlation analysis [15], analysis of variance, t-tests, multiple linear regression, continuous density hidden Markov models, Box-Jenkins transfer function models, time series analysis, and Mann-Whitney tests. In studies related to vaccination, the double difference model is commonly used. This model is a widely used econometric method in policy analysis and engineering evaluation, mainly applied to evaluate the impact of an event or policy in mixed cross-sectional datasets. For example, Díaz et al. used the "Difference-in-Difference-in-Differences" approach to estimate the impact of vaccination progress on the welfare of different socioeconomic groups, finding a positive correlation between mental health and the proportion of vaccinated individuals [16]. Recent studies have also developed AI-driven tools to predict and control epidemic waves using social media data. For instance, the EMIT (Epidemic and Media Impact Tool) model developed by researchers analyzes social media health communications to detect and predict pandemics. This model demonstrated its ability to predict future pandemic waves with high accuracy by integrating social media data trends, showcasing the potential of online data in informing public health interventions, including vaccination campaigns [17]. Additionally, mathematical models like those based on the extended SIR framework are being increasingly used in epidemiological studies, offering more precise predictions for airborne diseases, and could potentially complement internet search data models by simulating pandemic spread and the effect of interventions [18]. It can be seen that previous research methods have focused mainly on modeling and analyzing online data, with little

attention given to the interaction between online data and offline entities.

Offline: vaccine administration

According to statistics from Our World in Data at the University of Oxford, as of March 2024, 70.6% of the global population has received at least one dose of the COVID-19 vaccine. In low-income countries, 32.7% of the population has received at least one dose of the COVID-19 vaccine. Globally, over 13.5 billion doses of COVID-19 vaccine have been administered, with the vaccination rate still increasing at a pace of 10,383 doses per day. The top three countries in terms of cumulative vaccine doses administered are China (3.49 billion), India (2.21 billion), and the United States (670 million). In terms of the proportion of vaccinated population, approximately one-third of countries globally have achieved a vaccination rate of 75% or higher. Despite the increasing global COVID-19 vaccination rates, nearly one-third of the global population, particularly in low- and middle-income regions, has yet to be vaccinated. The African region has the lowest COVID-19 vaccination rate with at least one dose administered [19].

From the perspective of perceived risks of vaccination and how it affects individual interest in vaccination, extensive research has been conducted to understand the factors contributing to changes in overall vaccination rates. Studies have shown that factors such as direct and indirect costs of vaccination [20], implementation of vaccination campaigns [21], misinformation, and information scarcity [22] can all impact vaccination uptake. Different studies indicate that for some individuals, the intention to vaccinate increases during pandemics, while for others, vaccine hesitancy increases. This is an area that requires further investigation to understand the differences between these reactions and how to best encourage vaccine acceptance. For example, from January 2020 to September 2020, COVID-19 vaccine acceptance rates in Europe decreased from 70% to less than 50% [23]. In 2020, those opposed to receiving new vaccines primarily fell into two categories: those who mistrusted the government and those who believed that expediting vaccine development was unsafe and unpredictable [24]. However, due to concerns about vaccine safety and efficacy as well as general risk perception, acceptance rates decreased for both groups. In efforts to increase vaccination rates, methods such as correcting misinformation through social media [25], effectively utilizing government initiatives [26], and enhancing healthcare workers' training have been explored [27]. In the context of the internet's increasing influence on offline society, promoting dialogue, evoking emotions, and conducting propaganda through online searches are noteworthy directions

Liu *et al. BMC Public Health*   (2025) 25:975

Page 4 of 18

for increasing vaccination rates. Additionally, in the internet era, people tend to prefer actively searching for information online and proactively getting vaccinated.

In general, this study has made new advances compared to existing research in several aspects: (1) Proposing a method for detecting nonlinear correlations between online data and offline entities. (2) Building a new mathematical model based on the Logistic model to explicitly describe and quantify the interaction between internet searches and vaccine administration. (3) Through predictive research, this study provides a new method for forecasting the number of COVID-19 vaccine doses administered. This not only provides theoretical support for adjusting public health policies related to COVID-19 vaccine administration but also offers valuable insights for the administration of other vaccines.

Data modeling environment

Network data reflects social entities online and serves as a "barometer" of societal activities. Without social entities, network data would cease to exist. Following every instance of online public opinion, there lies an attribution to a social entity issue. However, research on the regulatory mechanisms governing the correlation between network data and social entities within the social system is relatively scarce after sudden events occur. There are two reasons for this: firstly, it is often challenging to quantify both network data and social entities simultaneously after a sudden event; secondly, the impact of some sudden events may be limited in time or space, rendering the quantified data less representative. So, how can we explore the regulatory mechanisms governing the correlation between network data and social entities within the social system?

From the perspective of sudden events, since the World Health Organization declared the outbreak of novel coronavirus pneumonia as an international public health emergency, the COVID-19 pandemic has persisted for nearly four years. Up to now, the novel coronavirus is still causing deaths and mutations. The COVID-19 pandemic, as a sudden public health emergency, has a long time span, a wide spatial scope, and is representative, making it suitable for studying the regulatory mechanisms governing the correlation between network data and social entities within the social system. Therefore, in the face of the big data environment triggered by the COVID-19 pandemic, how can we select appropriate methods for quantifying network data and social entities?

Quantitative data

As the most severe global public health emergency in the past century, the COVID-19 pandemic has brought about significant changes in people's lifestyles and survival status both online and offline. Online, social media has almost overnight become the primary means of communication and social interaction for people worldwide, with the quantity of public opinion information showing a high correlation with confirmed COVID-19 cases. The real-time dynamics of the pandemic have inundated media platforms, with daily updates on confirmed cases, infected regions, trajectory tracking, and incidents of mask hoarding constantly entering the public eye. Concerns among the public have shifted from when the pandemic will end to whether there will be effective vaccines to alleviate mortality. This high level of attention can affect the strength of people's awareness in adopting self-protection measures. Therefore, the public's long-term, widespread, and sustained attention to COVID-19 vaccines online meets the requirements for detecting the regulatory mechanisms governing the correlation between network data and social entities.

Baidu Index is a data analysis platform based on the behavioral data of Baidu's massive internet users (https://index.baidu.com/v2/index.html#/). It relies on daily search data from the Baidu platform, with keywords as statistical objects, to sum up the search volumes of various keywords on the Baidu platform weightedly. It is one of the most important statistical analysis platforms in the current era of the internet and data. Baidu Index consists of four attributes: Trend Research, Demand Map, Public Opinion Insight, and Audience Portrait. Due to different search interfaces, it is divided into PC and APP terminals. The development history of APP terminal search index is relatively short, recording statistical data only from 2011 to the present, while the PC terminal search index has a longer history, with relatively complete data from 2006 to the present. Among them, the Baidu Search Index is included in the Trend Research attribute. The search index is based on the search volume of netizens on Baidu, with keywords as statistical objects. It scientifically analyzes and calculates the weighted sum of the search frequency of each keyword in Baidu's web search. That is, the popularity of keywords searched by users in a unit of time can reflect the level of internet users' attention to keyword searches and their continuous changes. In this study, we crawled Baidu search index data for the keyword "COVID-19 vaccine" from March 23, 2021, to December 23, 2022, as the quantified data for network data.

In the offline realm, the public faces an unprecedented "new normal" in epidemic prevention and control. Various key groups such as medical workers, police officers, teachers, and students have had their work and lifestyle drastically altered. Since the release of the SARS-CoV-2 genome sequence in January 2020, researchers worldwide have been accelerating the development of vaccines

Liu *et al. BMC Public Health*   (2025) 25:975

Page 5 of 18

against COVID-19, resulting in the approval of multiple COVID-19 vaccines for emergency use by December 2020. Despite ongoing global efforts, the pandemic persists, with various mutated viruses emerging, suggesting that humanity may have to coexist with the COVID-19 virus for an extended period. Vaccination, aimed at enhancing immunity, may become a long-term process. In response to the complex and severe impact of the COVID-19 pandemic, governments worldwide have prioritized the development and administration of COVID-19 vaccines as a key measure to curb the outbreak [28]. Given the continuous mutation of the virus, future vaccination efforts may be necessary to strengthen immunity against different strains, making vaccination an enduring process. Therefore, the COVID-19 vaccine administration exhibits characteristics of long-term, widespread, and sustained impact, meeting the requirements for detecting the relational impact patterns in social entities.

In the era of the pandemic, COVID-19 vaccination has continuously transitioned towards public participation, consolidation, and coverage. Starting from March 23, 2021, the National Health Commission of China (http://www.nhc.gov.cn) began publishing the COVID-19 vaccination situation in China and updating the cumulative vaccination data daily (http://www.nhc.gov.cn/xcs/xxg-zbd/gzbd_index.shtml). In view of this, in this study, we crawled daily COVID-19 vaccination data in China from March 23, 2021, to December 23, 2022, quantified as the number of vaccinations administered per day (unit: 10,000 doses), as the quantitative data for social entities.

Data correlation analysis

After obtaining the initial data, analysis reveals two corresponding relationships in terms of time and data volume between Baidu search index data (SI) and China's COVID-19 vaccine inoculation data (VI). Firstly, there is a variability relationship where each data point in the BI corresponds to the search index quantity for each day. This variability relationship slices time and analyzes the data generated at each time point, corresponding to an instantaneous relationship. Secondly, there is an evolutionary relationship in the VI, where each data point corresponds to the cumulative number of vaccine inoculations up to that day. The evolutionary relationship accumulates time and analyzes all generated data, corresponding to an accumulated relationship. When exploring the rules of the association between network data and social entities in the social system, it is necessary to choose whether to use instantaneous data or cumulative data. Firstly, the data format needs to be standardized by segmenting VI into daily instantaneous data to obtain an instantaneous dataset; and SI is accumulated daily to obtain a cumulative dataset.

To investigate the rationality of data selection, we conducted correlation analysis separately on instantaneous data and cumulative data to determine the correlation between BI and VI. We identified the dataset with the highest correlation coefficient to be used for subsequent modeling work.

Correlation analysis is divided into two types: global correlation analysis and dynamic correlation analysis.

The global correlation coefficient is used to measure the overall association between BI and VI over the entire study period from March 24, 2021, to December 23, 2022, comprising 640 data points. We calculate the global correlation coefficient using the following formula:

$$r_{global} = cor(X, Y) = \frac{cov(X, Y)}{\sigma X \sigma Y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

Where r_{global} represents the global correlation coefficient and $cov(X, Y)$ represents the covariance between BI X and VI Y. σX and σY respectively denote the standard deviations of the search index X and COVID-19 vaccination data Y. The linear correlation coefficient eliminates the influence of dimensions. Its values range from -1 to 1, where a value closer to 1 indicates a stronger linear correlation, while a value of 0 indicates no linear correlation.

The calculation results are as follows:

$$r_{global1} = 0.782171172$$

$$r_{global2} = 0.992609001$$

Where $r_{global1}$ represents the global correlation coefficient between instantaneous data, and $r_{global2}$ represents the global correlation coefficient between cumulative data. $0.5 < r_{global} < 1$, indicating a strong positive correlation. It is evident that both X and Y exhibit strong correlation in both instantaneous and cumulative data. $r_{global2} > r_{global1}$, indicating that the correlation in cumulative data is stronger compared to instantaneous data.

The dynamic correlation coefficient is used to describe the dynamic changes in the correlation between X and Y from the beginning to the t-th day. By calculating the evolution correlation coefficient $r_{evolution}(t)$, we can observe the dynamic changes in the correlation between X and Y during the evolutionary process. Its calculation formula is as follows:

$$r_{evolution}(t) = \frac{cov(X_{1:t}, Y_{1:t})}{\sigma X_{1:t} \sigma Y_{1:t}} = \frac{\sum_{i=1}^{t}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{t}(x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{t}(y_i - \bar{y})^2}}$$

Where $r_{evolution}(t)$ represents the dynamic correlation coefficient for the t-th day, $cov(X_{1:t}, Y_{1:t})$ denotes the

Liu *et al. BMC Public Health*   (2025) 25:975

Page 6 of 18

covariance between X and Y from the beginning of the event to the t-th day, and $\sigma X_{1:t}$ and $\sigma Y_{1:t}$ respectively represent the standard deviations of X and Y during that time period. Since the Pearson correlation coefficient requires at least two data points to analyze the correlation, in the analysis of dynamic correlation coefficients, a total of 639 Pearson correlation coefficients were obtained. The dynamic trend of the correlation coefficient over time is shown in Fig. 1.

It can be observed that the dynamic correlation coefficient $r_{evolution1}$ of instantaneous data fluctuates more significantly compared to the average dynamic correlation coefficient $avg(r_{evolution1}(t)) = 0.703846124$. The dynamic correlation coefficient $r_{evolution2}$ of cumulative data exhibits smaller fluctuations, with the lowest value $r_{evolution2}(53)$ observed at 53 days, where $r_{evolution2}(53) = 0.975521704$. The average dynamic correlation coefficient $avg(r_{evolution2}(t)) = 0.99172064$ indicates that the overall dynamic correlation coefficient of cumulative data remains at a very high level.

From the results of the overall and dynamic correlation coefficients, it can be observed that the cumulative data of BI and VI exhibit higher correlation compared to the instantaneous data, indicating greater research value and significance. Furthermore, this also confirms the rationality of our data quantification and fundamental assumptions regarding network data and social entities, providing support for our subsequent establishment of basic models. In terms of the meaning of the data itself, cumulative data considers historical factors and past influences, while instantaneous data focuses more on the current state. We need to analyze historical evolutionary patterns and capture trends in data evolution over time.

Therefore, we choose cumulative data as the final modeling data, with a time range from March 24, 2021, to December 23, 2022, totaling 640 data points.

Model establishment

Hypothesis

The evolution pattern of information data generated by sudden events has been extensively studied, although the models applied and the delineation of stages may vary. Nevertheless, fundamentally, they all acknowledge that information data evolve along an "S-curve". In 1985, the renowned American information resource management expert, Horton, introduced the concept of information lifecycle, suggesting that information resources follow natural laws of motion and possess their own lifecycle. According to the information lifecycle theory, the evolution process of information data can be divided into stages. Currently, the academic community has different views on the division and naming of stages in the evolution of information lifecycle, which can range from three to six stages depending on different criteria. However, overall, different stage divisions follow the evolutionary pattern of information generation, development, and extinction. It has been observed that the evolution of Baidu search index and China's COVID-19 vaccine administration also exhibits characteristics of an incubation period, a diffusion period, and a decline period.

Therefore, the logistic model proposed by the Belgian mathematician P.F. Verhulst, initially used to study the growth of social populations, can be employed to simulate the evolution process of information data generated by sudden events in both online and offline domains. It is assumed that the quantified data of
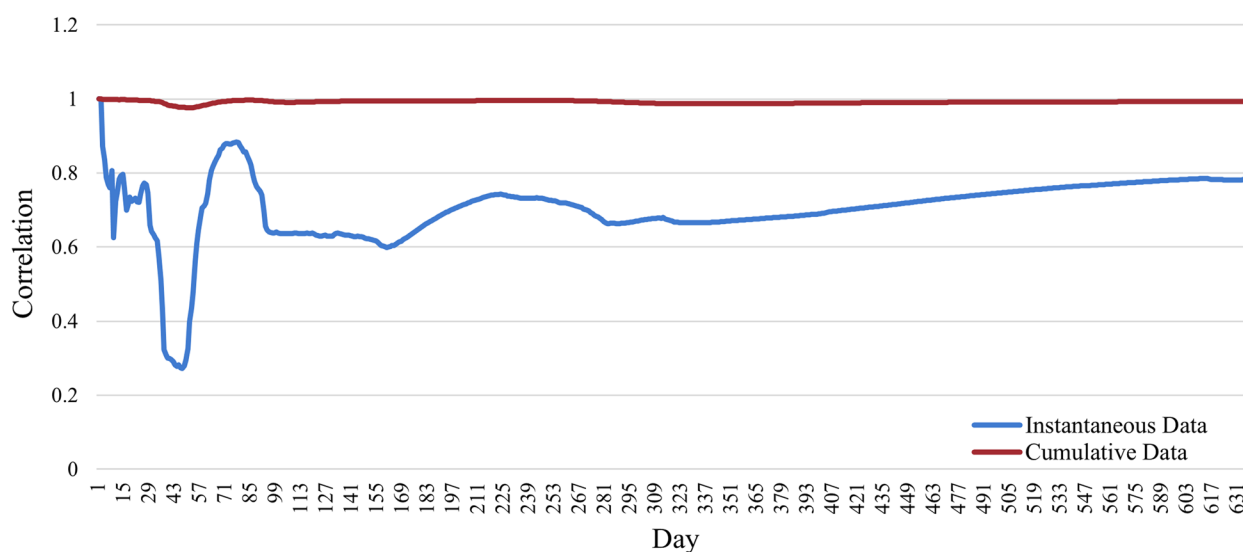

Fig. 1 Dynamic trend of Pearson correlation coefficient over time

Liu *et al. BMC Public Health*   (2025) 25:975

Page 7 of 18

online and offline entities conform to an independently existing logistic model.

Specifically, it is assumed that the BI $x_1(t)$, corresponding to online data, is a monotonically increasing function of time t with an initial value of $x_1(0)$. The number of VI $x_2(t)$, corresponding to social entities, is also a monotonically increasing function of time t with an initial value of $x_2(0)$. Based on the above analysis, the foundational model for obtaining quantified data of online and offline entities (model 1) is constructed as follows:

$$\begin{cases} \frac{dx_1}{dt} = r_1 x_1 \left(1 - \frac{x_1}{K_1}\right) \\ \frac{dx_2}{dt} = r_2 x_2 \left(1 - \frac{x_2}{K_2}\right) \end{cases} \tag{1}$$

Where $r_1 > 0$ represents the inherent growth rate of the cumulative BI, and $r_2 > 0$ represents the inherent growth rate of cumulative VI. Due to the existence of a lifecycle, both the cumulative BI and cumulative VI have upper limits. Here, K_1 represents the upper limit of the cumulative BI, and K_2 represents the upper limit of cumulative VI. Therefore, the remaining space for the cumulative BI is denoted as $\left(1 - \frac{x_1}{K_1}\right)$, and the remaining space for cumulative VI is denoted as $\left(1 - \frac{x_2}{K_2}\right)$.

To verify the rationality and accuracy of the above assumptions, we conducted a fitting analysis on the full-cycle data of the basic model. First, it is necessary to convert the basic model into the corresponding difference equations. After the transformation, the parameter fitting problem of the differential equations is converted into a regression analysis problem of the difference equations. The specific system of equations is as follows:

$$\begin{cases} \Delta x_1(t) = r_1 x_1 \left(1 - \frac{x_1}{K_1}\right) = r_1 x_1 - \frac{r_1}{K_1} x_1{}^2 \\ \Delta x_2(t) = r_2 x_2 \left(1 - \frac{x_2}{K_2}\right) = r_2 x_2 - \frac{r_2}{K_2} x_2{}^2 \end{cases} \tag{2}$$

Where $\Delta x_1(t) = x_1(t) - x_1(t-1)$, $\Delta x_2(t) = x_2(t) - x_2(t-1)$, $\Delta x_1(t)$ and $\Delta x_2(t)$ respectively represent the statistical quantities of BI and VI at time t, while $x_1(t)$ and $x_2(t)$ represent the cumulative quantities of BI and VI at time t. During data processing, it was found that, due to the need to calculate $\Delta x(t)$, the data starts from the second item and has numerical values, resulting in a total of 640 data points. Observing the equations, it can be noted that the difference $\Delta x_1(t)$ exhibits a bivariate linear relationship with x_1 and $x_1{}^2$, and the same applies to $\Delta x_2(t)$.

Next, using bivariate linear regression analysis, a regression analysis is performed on each of the two variables in this dataset separately, yielding regression coefficients r_1, $-\frac{r_1}{K_1}$, r_2, and $-\frac{r_2}{K_2}$. This allows us to determine the parameters r_1, K_1, r_2, and K_2 of the aforementioned differential equations.

By observing the model 1, it is noted that in practical scenarios, the intrinsic growth rates r_1 and r_2, as well as the quantity limits K_1 and K_2, all need to be greater than 0. Hence, $-\frac{r_1}{K_1} < 0$ and $-\frac{r_2}{K_2} < 0$. Therefore, this can be set as a structural test criterion.

In addition, to make the results more accurate and representative, we also introduced goodness-of-fit and significance tests. The goodness-of-fit test introduces the statistical measure R-squared, which is a measure of the goodness of fit of a linear regression model. Typically, the larger the R-squared, the better the regression model fits the observed data. We set a criterion for goodness-of-fit test to pass when R-squared is greater than 0.36. The significance test introduces the *P*-value. We set the criterion for passing the significance test to be when the *P*-value is less than 0.05.

After conducting linear regression analysis on BI and VI, the results were examined using the three aforementioned testing methods. Equation 1 represents the condition where $\frac{dx_1}{dt} = r_1 x_1 \left(1 - \frac{x_1}{K_1}\right)$, and equation 2 represents the condition where $\frac{dx_2}{dt} = r_2 x_2 \left(1 - \frac{x_2}{K_2}\right)$. The results are summarized in Table 1.

The results indicate that in Equation 1, both parameters pass the structural test for signs, with $R^2 > 0.36$ meeting the goodness of fit test. The *P*-values for both coefficients are less than 0.005, indicating strong significance in the regression coefficients, satisfying the significance test. In Equation 2, both parameters pass the structural test for signs, with $R^2 > 0.36$ meeting the goodness of fit test. The P-values for both coefficients are less than 0.005, indicating strong significance in the regression coefficients. Both equations pass the triple test, suggesting a high degree of fit between the data and the basic differential equations. Therefore, we can conclude that the above assumptions are reasonable, allowing us to proceed with the next data modeling task.

Data modeling

The preceding research has confirmed that there is a strong correlation between BI and VI, serving as

Table 1 Results Verification

	r_i $(i = 1, 2)$	r_i $P - value$	$-\frac{r_i}{K_i}$	$-\frac{r_i}{K_i}$ $P - value$	R^2
Equation 1	0.013015	1.3E − 105	−9.35E − 10	2.34E − 92	0.581097
Equation 2	0.014121	4.1E − 158	−4.12E − 08	1.6E − 143	0.709028

Liu *et al. BMC Public Health*   (2025) 25:975

Page 8 of 18

quantified data for network data and social entities. Following the outbreak of the COVID-19 pandemic, the public's attention to effective vaccines and the urgency of protecting the population have driven the development, production, and deployment of COVID-19 vaccines. The administration of COVID-19 vaccines, in turn, influences public attention. However, how do internet searches and vaccine administration interact with each other? What effects do they generate? What are the mechanisms behind these effects? Further exploration is still needed to address these questions.

In order to explore the underlying mechanisms of the association between online searches and vaccination, we further proceed with data modeling. We use function f_1 to describe the impact of the cumulative VI on the growth rate of cumulative BI. Function f_2 is employed to depict the influence of the cumulative BI on the growth rate of cumulative VI. Let f_1 and f_1 denote the influence functions, representing the coexistence of online search data and societal entities on the fundamental assumptions of the model. The detection model for the mechanism of correlated influence (model 3) can be represented in the following form:

$$\begin{cases} \frac{dx_1}{dt} = r_1 x_1 \left(1 - \frac{x_1}{K_1}\right) + f_1(x_1, x_2) \\ \frac{dx_2}{dt} = r_2 x_2 \left(1 - \frac{x_2}{K_2}\right) + f_2(x_1, x_2) \end{cases} \qquad (3)$$

Firstly, we need to determine the structure of the two influence functions, f_1 and f_2. As known from the aforementioned, in the basic Logistic model, $\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right) = rx - \frac{r}{K}x^2$, and the rate of change $\frac{dx}{dt}$ is a binary linear function of x and x^2, where the highest order of variables is quadratic, and there is no constant term. Therefore, we assume that when quantifying the correlation between online search data and societal entities, the highest order of a single variable is quadratic. Thus, taking influence function $f_1(x_1, x_2)$ as an example, it can include at most six structural factors: x_2, x_2^2, $x_1 x_2$, $x_1 x_2^2$, $x_1^2 x_2$, $x_1^2 x_2^2$. These six structural factors can be combined to form terms containing i structural factors ($i = 1, 2, 3, 4, 5, 6$). There are a total of $C_6^1 + C_6^2 + C_6^3 + C_6^4 + C_6^5 + C_6^6 = 63$ possibilities for the factors in $f_1(x_1, x_2)$. Additionally, the signs preceding these six structural factors can be either positive or negative. Similarly, $f_2(x_1, x_2)$ can also include at most six structural factors: x_1, x_1^2, $x_1 x_2$, $x_1 x_2^2$, $x_1^2 x_2$, $x_1^2 x_2^2$, resulting in 63 possible structures.

At this point, the detection process for the associative impact pattern mechanism between network data and social entities has been established, aiming to filter out the common structural factors that meet the model requirements. The specific process is as follows:

(1) Construct the model formula and convert it into a set of difference equations. Transform model 3 into corresponding difference equations, forming two sets of difference equations collections S_1 and S_2, each containing 63 equations.

$$\begin{cases} \Delta x_1(t) = r_1 x_1 \left(1 - \frac{x_1}{K_1}\right) + f_1(x_1, x_2) \\ \Delta x_2(t) = r_2 x_2 \left(1 - \frac{x_2}{K_2}\right) + f_2(x_1, x_2) \end{cases} \qquad (4)$$

Where $\Delta x_1(t) = x_1(t) - x_1(t-1)$, $\Delta x_2(t) = x_2(t) - x_2(t-1)$, $\Delta x_1(t)$ and $\Delta x_2(t)$ represent the corresponding BI and VI in time t ($t = 1,2,3,\ldots,640$).

(2) Conduct regression analysis. We perform two types of regression analysis: static regression analysis based on global data and dynamic regression analysis based on evolving data. Static regression analysis refers to evaluating whether there is a strong correlation between BI and VI from March 24, 2021, to December 23, 2022, covering the entire period. Dynamic regression analysis involves starting with a subset of data and adding one data point at a time for regression analysis until the entire period is covered.

(3) We perform a triple regression test. The triple test includes structural, correlation, and significance tests. In the structural test, the first parameter of equation $\Delta x(n) = rx\left(1 - \frac{x}{K}\right) + f(x_1, x_2) = rx - \frac{r}{K}x^2 + f(x_1, x_2)$ needs to be positive, and the second parameter needs to be negative, satisfying the structural assumptions of the model. In the correlation test, the change rates $\frac{dx}{dt}$ and $\frac{dy}{dt}$ have a strong correlation with the influencing factors, with a criterion of passing the fit test when R-squared is greater than 0.36. In the significance test, the significance of the resulting parameters is measured, passing the significance test when the P-value is less than 0.05.

(4) Identify common structures and quantify the impact patterns. Among the 63 structural factors mentioned above, we select based on the highest dynamic test passing rate, obtaining the structure of the impact function with the greatest commonality. Here, we introduce the Passing Rate (PR) to describe the proportion of times a structural factor passes the triple regression test out of the total number of times that the structural factor needs to undergo dynamic regression analysis. The structural factor with the highest passing rate best reflects representative patterns of correlated impact.

$$PR = \frac{N_i}{N_{total}}$$

In this context, N_{total} represents the number of times each structural equation needs to undergo dynamic regression analysis, which is 621 for the selected data. N_i

Liu *et al. BMC Public Health*   (2025) 25:975

Page 9 of 18

Table 2 PR results

Structure factor	Passing count	Passing rate	Structure factor	Passing count	Passing rate
x_2	58	0.093397746	$-x_2$	310	0.499194847
x_2^2	374	0.602254428	$-x_2^2$	0	0
x_1x_2	351	0.565217391	$-x_1x_2$	0	0
$x_1x_2^2$	564	0.90821256	$-x_1x_2^2$	0	0
$x_1^2x_2$	564	0.90821256	$-x_1^2x_2$	0	0
$x_1^2x_2^2$	563	0.906602254	$-x_1^2x_2^2$	1	0.001610306

represents the number of times a particular structural factor passes the triple regression test.

Data modeling for internet searches

We initiated the analysis from the first 20 data points (covering from March 24, 2021, to April 12, 2021), incrementally adding one data point at a time until reaching the data points of December 23, 2022. Each structural equation (totaling 63) underwent 621 dynamic regression analyses, aiming to evaluate the stability of the detected associative impact pattern mechanisms during the evolutionary process.

We divided the 63 structural factors into groups based on the number of terms, ranging from one to six terms (where one term contains C_6^1 structural factors, two terms contain C_6^2 structural factors, and so on, up to six terms containing C_6^6 structural factors). After regression analysis, we obtained the values of the Passing Rate (PR) for each structural factor.

The PR results for the first equation's one term are shown in Table 2.

From the table, it can be observed that when the number of structural factors is one, structural factors $x_1x_2^2$ and $x_1^2x_2$ have the highest number of passes in the threefold regression out of 621 dynamic regressions, with a passing rate as high as 90.82%.

Extracting the structural factors with the highest passing rate for each item (for example, when the number of structural factors is one, extracting $x_1x_2^2$ and $x_1^2x_2$), we obtain Table 3.

Overall, when the structural factors are $x_1x_2^2$ and $x_1^2x_2$, the pass rate is the highest at 90.82%. That is, the impact function of VI on BI is $f_{11}(x_1, x_2) = a_1x_1x_2^2$, $a_1 > 0$, where a_1 is the coefficient of structural factor $x_1x_2^2$; or $f_{12}(x_1, x_2) = a_2x_1^2x_2$, $a_2 > 0$, where a_2 is the coefficient of structural factor $x_1^2x_2$.

Performing a global static regression analysis on the obtained differential equation $\Delta x_{11}(t) = r_{11}x_1\left(1 - \frac{x_1}{K_{11}}\right) + a_1x_1x_2^2 = r_{11}x_1 - \frac{r_{11}}{K_{11}}x_1^2 + a_1x_1x_2^2$ yields the fitting results, as shown in Table 4.

Table 3 Structural factors with the highest passing rate for each item

	Structure factor	Passing rate(%)
One Item	$x_1x_2^2$ or $x_1^2x_2$	90.82
Two Item	$-x_2^2 + x_1^2x_2$	85.99
Three Item	$-x_2^2 + x_1x_2^2 - x_1^2x_2^2$	59.58
Four Item	$-x_2 + x_1x_2 - x_1x_2^2 + x_1^2x_2^2$	12.72
Five Item	$-x_2^2 + x_1x_2 + x_1x_2^2 - x_1^2x_2 - x_1^2x_2^2$	47.02
Six Item	$-x_2 - x_2^2 + x_1x_2 - x_1x_2^2 - x_1^2x_2 - x_1^2x_2^2$	3.54

Table 4 Fitting results

Parameters	Values	P-value	Whether it has passed the test
r_{11}	0.024136	5.9E − 95	Passed
$-\frac{r_{11}}{K_{11}}$	−2.9E − 09	1.25E − 60	Passed
a_1	1.44E − 13	2.81E − 33	Passed
R_1^2	0.663705	——	Passed

After conducting full-cycle static regression analysis, the equation $\frac{dx_1}{dt} = 0.024136x_1 + (-2.9E - 09)x_1^2 + (1.44E - 13)x_1x_2^2$ was obtained, and the fitting results passed the triple test, with all P-values significantly less than 0.005, indicating a strong significance of the regression coefficients.

Performing global static regression analysis on the obtained differential equation $\Delta x_{12}(t) = r_{12}x_1\left(1 - \frac{x_1}{K_{12}}\right) + a_2x_1^2x_2 = r_{12}x_1 - \frac{r_{12}}{K_{12}}x_1^2 + a_2x_1^2x_2$ yields the fitting results, as shown in Table 5.

After the full-period static regression analysis, the equation $\frac{dx_1}{dt} = 0.037413x_1 + (-5.4E - 09)x_1^2 + (8.18E - 15)x_1^2x_2$ was obtained, and the fitting results passed the triple test, with P-values far below 0.005, indicating a very strong significance of the regression coefficients. Comparing Tables 4 and 5, when the structure factors are $x_1x_2^2$ and $x_1^2x_2$, $r_{1i}(i = 1, 2) > 0$, $-\frac{r_{1i}}{K_{1i}} < 0$, satisfying the structural test. In terms of the goodness of fit test, $R_2^2 > R_1^2$, indicating that the equation with the structure factor $x_1^2x_2$ has

Liu *et al. BMC Public Health*   (2025) 25:975

Page 10 of 18

Table 5 Fitting results

Parameters	Values	$P-$value	Whether it has passed the test
r_{12}	0.037413	$1.1E-146$	Passed
$-\frac{r_{12}}{K_{12}}$	$-5.4E-09$	$1.1E-114$	Passed
a_2	$8.18E-15$	$2.44E-89$	Passed
R_2^2	0.775475	——	Passed

a better fit. In terms of significance test, the P-values of parameters r_{12}, $-\frac{r_{12}}{K_{12}}$, and a_2 are smaller compared to the P-values of parameters r_{11}, $-\frac{r_{11}}{K_{11}}$, and a_1, indicating that the equation with the structure factor $x_1^2 x_2$ has better significance. Therefore, the structure factor $x_1^2 x_2$ is selected.

Data modeling for COVID-19 vaccine administration
Similar to the data modeling of web searches, after conducting regression analysis on $\triangle x_2(t) = r_2 x_2 \left(1 - \frac{x_2}{K_2}\right) + f_2(x_1, x_2)$, taking the number of structural factors as one item as an example, the PR results are shown in Table 6.

From the table, it can be seen that when the number of structural factors is one, the structural factor $x_2^2 x_1$ has the highest passing rate, with a rate as high as 89.37% in 621 dynamic regression tests.

Extracting the structural factor with the highest passing rate for each item (for example, when the number of structural factors is one, extracting $x_2^2 x_1$), Table 7 is obtained.

Overall, when the structural factor is $x_2^2 x_1$, the passing rate is the highest at 89.37%. That is, the impact function of BI on VI is $f_2(x_1, x_2) = b x_2^2 x_1$, $b > 0$, where b is the coefficient of structural factor $x_2^2 x_1$.

For the obtained difference equation $\triangle x_2(t) = r_2 x_2 \left(1 - \frac{x_2}{K_2}\right)$ $+ b x_2^2 x_1 = r_2 x_2 - \frac{r_2}{K_2} x_2{}^2 + b x_2^2 x_1$, conducting a global static regression analysis yields the fitted results, as shown in Table 8.

Through the static regression analysis of the entire period, the equation $\frac{dx_2}{dt} = 0.019741 x_2 + (-9.2E-08) x_2{}^2 + (2.45E-15) x_2^2 x_1$ is obtained, and the fitted result has passed the triple test, with

all P-values significantly below 0.005, indicating strong significance of the regression coefficients.

Model comparison
Compare the fitting results obtained from the data modeling with those of the basic model. Since all fitting results have passed the structural test, in the significance test, the optimized model has one more parameter compared to the basic model, making it impossible to perform a comparative analysis. Therefore, selecting the parameter R^2 from the goodness-of-fit test as the standard for comparative analysis, as shown in Table 9.

It can be observed that after adding the optimal structural factors, the fitness of the model has significantly improved, demonstrating the effectiveness and rationality of the data modeling process.

Model analysis
Through quantifying network data and social entities, we have constructed a model for detecting associative impact patterns between network data and social entities. We have introduced a triple test method and obtained the optimal associative impact pattern common structural equation with the highest fitness. Given that the identified optimal structural factors exhibit positive values, this study infers a positive interaction between online search behavior and vaccination rates, which drives a mutually reinforcing relationship between the two. Based on this, the study terms this bidirectional promotion phenomenon as the "symbiotic effect" to highlight the mutually beneficial relationship between online search and vaccination in their dynamic interaction. In this section, we will explore and analyze the mechanisms implied by the model, and conduct a detailed analysis of the formation and implications of the symbiotic effects between internet searches and vaccine administration.

Model consolidation
The optimal model for the most significant correlation pattern between BI and VI, as explored in the previous section, is summarized as follows:

Table 6 PR results

Structure factor	Passing count	Passing rate	Structure factor	Passing count	Passing rate
x_1	0	0	$-x_1$	12	0.019323671
x_1^2	0	0	$-x_1^2$	369	0.594202899
$x_2 x_1$	0	0	$-x_2 x_1$	55	0.088566828
$x_2 x_1^2$	343	0.552334944	$-x_2 x_1^2$	0	0
$x_2^2 x_1$	555	0.893719807	$-x_2^2 x_1$	0	0
$x_2^2 x_1^2$	554	0.892109501	$-x_2^2 x_1^2$	6	0.009661836

Liu *et al. BMC Public Health*   (2025) 25:975

Page 11 of 18

Table 7 Structural factors with the highest passing rate for each item

	Structure factor	Passing rate(%)
One Item	$x_2^2 x_1$	89.37
Two Item	$-x_1^2 + x_2 x_1^2$	78.10
Three Item	$-x_1^2 + x_2^2 x_1 - x_2^2 x_1^2$	54.91
Four Item	$-x_2 x_1 + x_2 x_1^2 + x_2^2 x_1 - x_2^2 x_1^2$	23.35
Five Item	$-x_1^2 + x_2 x_1 - x_2 x_1^2 + x_2^2 x_1 - x_2^2 x_1^2$	39.77
Six Item	——	0

$$\begin{cases} \frac{dx_1}{dt} = r_1 x_1 \left(1 - \frac{x_1}{K_1}\right) + f_1(x_1, x_2) = r_1 x_1 - \frac{r_1}{K_1} x_1^2 + a_2 x_1^2 x_2 \\ \frac{dx_2}{dt} = r_2 x_2 \left(1 - \frac{x_2}{K_2}\right) + f_2(x_1, x_2) = r_2 x_2 - \frac{r_2}{K_2} x_2^2 + b x_2^2 x_1 \end{cases} \quad (5)$$

Breaking it down and simplifying, we get the system of equations:

$$\begin{cases} \frac{dx_1}{dt} = r_1 x_1 \left(1 - \frac{x_1}{K_1} + a_2 \frac{K_1 K_2}{r_1} \frac{x_2}{K_2} \frac{x_1}{K_1}\right) = r_1 x_1 \left(1 - \frac{x_1}{K_1} + \alpha \frac{x_2}{K_2} \frac{x_1}{K_1}\right) \\ \frac{dx_2}{dt} = r_2 x_2 \left(1 - \frac{x_2}{K_2} + b \frac{K_1 K_2}{r_2} \frac{x_1}{K_1} \frac{x_2}{K_2}\right) = r_2 x_2 \left(1 - \frac{x_2}{K_2} + \beta \frac{x_1}{K_1} \frac{x_2}{K_2}\right) \end{cases}$$
$$(6)$$

Where $\alpha = a_2 \frac{K_1 K_2}{r_1}$ and $\beta = b \frac{K_1 K_2}{r_2}$, with specific symbol meanings as shown in Table 10.

Mechanism analysis

In natural ecosystems, population growth follows the Logistic model due to limitations imposed by natural environmental resources. This pattern is characterized by a gradual decrease in growth rate as population density increases, starting with rapid growth in the early stages, slowing down after reaching a certain size, and eventually reaching saturation. Similarly, in the foundational model, Baidu search index and China's COVID-19 vaccine administration volume undergo an evolutionary process from formation, growth, to stabilization. However, in social systems, connections exhibit universality, with internal elements of phenomena and interactions between phenomena mutually influencing, constraining, and interacting with each other. After the occurrence of the COVID-19 pandemic, the public's high attention to effective vaccines affects the strength of individuals' awareness of self-protection measures. Furthermore, following the development and

Table 8 Fitted results

Parameters	Values	$P-\text{value}$	Whether it has passed the test
r_2	0.019741	2.36E − 95	Passed
$-\frac{r_2}{K_2}$	−9.2E − 08	5.24E − 41	Passed
b	2.45E − 15	1.11E − 14	Passed
R^2	0.712115	——	Passed

Table 9 Fitting comparison

	Baseline model's R^2	Optimized model's R^2
Equation 1	0.581097	0.775475
Equation 2	0.709028	0.712115

production of COVID-19 vaccines, they become a focal point of public attention, influencing public search behavior on the internet regarding COVID-19 vaccines.

For the BI equation $\frac{dx_1}{dt} = r_1 x_1 \left(1 - \frac{x_1}{K_1} + \alpha \frac{x_2}{K_2} \frac{x_1}{K_1}\right)$, COVID-19 vaccination x_2 promotes an increase in the Baidu search index x_1. Firstly, the production and application of COVID-19 vaccines lead to a surge in public demand for information, resulting in an increase in online search behavior. The degree to which a unit of VI promotes BI x_1 is represented by $\alpha \frac{x_1}{K_1}$, while $\frac{x_2}{K_2}$ represents the degree to which $\alpha \frac{x_2}{K_2} \frac{x_1}{K_1}$ promotes BI x_1, where α is the coefficient of influence effect. When $\alpha > 0$, it signifies a promotion effect coefficient. Secondly, VI x_2 leads to an increase in the residual space for growth of BI x_1, from $1 - \frac{x_1}{K_1}$ to $1 - \frac{x_1}{K_1} + \alpha \frac{x_2}{K_2} \frac{x_1}{K_1}$. Based on these two points, offline COVID-19 vaccination drives online Baidu search index changes, exerting a positive impact on the change rate of the Baidu search index.

For the VI equation $\frac{dx_2}{dt} = r_2 x_2 \left(1 - \frac{x_2}{K_2} + \beta \frac{x_1}{K_1} \frac{x_2}{K_2}\right)$, the BI x_1 promotes an increase in the VI x_2. Firstly, after the outbreak of the COVID-19 pandemic, media platforms are inundated with real-time pandemic updates, and various pandemic-related public opinion events continuously enter the public's view. The public gradually starts to care and actively search for whether there are any effective vaccines available to alleviate the fatalities. This heightened attention leads to an increased awareness among people to take self-protective measures. The reinforcement of this self-protective awareness is reflected in

Table 10 Symbol meanings

Variables and parameters	Meaning
t	Time
x_1	Cumulative BI
x_2	Cumulative VI
r_1	Inherent growth rate of cumulative BI
r_2	Inherent growth rate of cumulative VI
K_1	The upper limit of the cumulative BI
K_2	The upper limit of the cumulative VI
α	The coefficient representing the correlation effect of VI on BI
β	The coefficient representing the correlation effect of BI on VI

Liu *et al. BMC Public Health*   (2025) 25:975

Page 12 of 18

social entities, resulting in an increase in COVID-19 vaccination volume. The degree to which a unit of BI promotes the VI x_2 is represented by $\beta \frac{x_2}{K_2}$, while $\frac{x_1}{K_1}$ represents the degree to which $\beta \frac{x_1}{K_1} \frac{x_2}{K_2}$ promotes the VI x_2. Here, β is the coefficient of influence effect, and when $\beta > 0$, it signifies a promotion effect coefficient. Secondly, the BI x_1 leads to an increase in the residual space for growth of VI x_2, from $1 - \frac{x_2}{K_2}$ to $1 - \frac{x_2}{K_2} + \beta \frac{x_1}{K_1} \frac{x_2}{K_2}$. Based on these two points, online-related Baidu search index also serves as a driving force for offline COVID-19 vaccination, exerting a positive impact on the change rate of COVID-19 vaccination volume.

Analysis of Interaction processes

Through modeling analysis, we have determined that the associative impact pattern between BI and VI is a mutually reinforcing interactive structure, indicating a symbiotic relationship between internet searches and vaccine administration. Information and activities related to COVID-19 vaccine administration trigger public search demands, thereby driving up the cumulative BI. Conversely, accessing search results also helps the public better understand and engage in vaccine administration, further promoting an increase in VI. This bidirectional promotion mechanism is consistent with the mechanism described in ecological models of population symbiosis in mathematical ecology. This provides a theoretical basis for explaining the model mechanism and lends greater practical significance to our findings.

From a public perspective, interest in vaccine administration gradually increases in the early stages of COVID-19 vaccine rollout. As people seek information about vaccine administration, they utilize search engines, leading to an increase in the cumulative search index. After COVID-19 vaccine administration, recipients share their vaccination experiences and engage in discussions on various forums and social media platforms. These shares and discussions generate more content related to searches, further driving up the cumulative search index on Baidu. For those who have been vaccinated, they may search for information regarding the efficacy and side effects of the vaccine, confirming the effectiveness of vaccination. These search behaviors also contribute to the increase in the BI. From a media perspective, as a highly discussed topic, the media reports and promotes vaccine administration, triggering public search demands and subsequently increasing the BI. From a governmental standpoint, the government continuously updates and adjusts vaccine administration policies and arrangements. The public searches for relevant information through search engines and makes appointments for vaccine administration accordingly. This series of actions also contributes to the increase in the BI.

For vaccine administration, an increase in BI implies a gradual rise in public attention towards COVID-19 vaccine administration. As information spreads, more people become aware of the importance of vaccine administration, and positive public discourse further drives the popularization of vaccination. These positive messages include the benefits, safety, and post-vaccination effects, which help dispel public doubts and increase willingness and acceptance of vaccination. Therefore, online searches can be utilized to promote vaccine-related campaigns [29] or enhance healthcare workers' education [30]. A survey conducted in Melbourne, Australia, indicated that online searches are an effective means to boost participation and coverage rates in influenza vaccination campaigns. Additionally, online search platforms can be used to provide the public with the latest, unbiased, and trustworthy information. For example, state health departments in the United States utilize Facebook to disseminate vaccine-related information, which constitutes 7% of all content published [31].

Through active engagement, the public can also gain access to information regarding vaccine appointment scheduling, vaccination locations, and other related details. This information helps the public understand how to go about getting vaccinated, thereby enhancing the accessibility of vaccine administration. From a policy perspective, during the vaccination period, the public pays closer attention to changes in vaccination policies, such as the expansion of age groups eligible for vaccination and increases in vaccine supply, all of which contribute to improving vaccination coverage rates.

Model application

By seeking the optimal common structural factors, we have derived a model describing the associative impact patterns between BI and VI. However, the greater value of the model lies in its application. Therefore, we will further explore the specific application value of this associative impact pattern model.

The value of big data lies in prediction, and the key to detecting associative impact patterns is how to utilize existing data to forecast future data trends. In practical applications, it is essential to first determine the parameters of the model in order to identify the types of equilibrium points and subsequently investigate future evolutionary trends.

First, determine the parameters of the model using the current BI and VI.

Let $\begin{cases} \frac{dx_1}{dt} = r_1 x_1 \left(1 - \frac{x_1}{K_1} + \alpha \frac{x_2}{K_2} \frac{x_1}{K_1}\right) = 0 \\ \frac{dx_2}{dt} = r_2 x_2 \left(1 - \frac{x_2}{K_2} + \beta \frac{x_1}{K_1} \frac{x_2}{K_2}\right) = 0 \end{cases}$, the equilibrium points of the model are as follows:

$$\begin{cases} P_0(0,0) \\ P_1\left(\frac{K_1}{2\beta}\left[1+\beta-\alpha+\sqrt{(\alpha-\beta)^2-2\alpha-2\beta+1}\right], \frac{K_2}{2\alpha}\left[1+\alpha-\beta+\sqrt{(\alpha-\beta)^2-2\alpha-2\beta+1}\right]\right) \\ P_2\left(\frac{K_1}{2\beta}\left[1+\beta-\alpha-\sqrt{(\alpha-\beta)^2-2\alpha-2\beta+1}\right], \frac{K_2}{2\alpha}\left[1+\alpha-\beta-\sqrt{(\alpha-\beta)^2-2\alpha-2\beta+1}\right]\right) \\ P_3(K_1,0) \\ P_4(0,K_2) \end{cases} \tag{7}$$

According to the theory: The local equilibrium point is stable when the determinant $DET(J) > 0$ and the trace $Tr(J) < 0$ of the Jacobian matrix. We could screen out the equilibrium points that meet the conditions by substituting each equilibrium point into the Jacobian matrix, as shown in Table 11.

According to the definition in this text, equilibrium points need to be greater than zero. P1 is not the equilibrium point sought in this paper. Only the case where P2 holds is discussed in this paper.

According to the model, the equilibrium point for the BI changes from K_1 to $\frac{K_1}{2\beta}\left[1+\beta-\alpha-\sqrt{(\alpha-\beta)^2-2\alpha-2\beta+1}\right]$, while the equilibrium point for the VI changes from K_2 to $\frac{K_2}{2\alpha}\left[1+\alpha-\beta-\sqrt{(\alpha-\beta)^2-2\alpha-2\beta+1}\right]$. Therefore, based on the previous n sets of BI and VI, dynamic prediction can be performed by adding new data to calculate the new equilibrium points and predict future trends [32]. Let $X(t)$ represent the true data at time t, $X_{balance1}$ represent the equilibrium point $\frac{K_1}{2\beta}\left[1+\beta-\alpha-\sqrt{(\alpha-\beta)^2-2\alpha-2\beta+1}\right]$, $X_{balance2}$ represent the equilibrium point $\frac{K_2}{2\alpha}\left[1+\alpha-\beta-\sqrt{(\alpha-\beta)^2-2\alpha-2\beta+1}\right]$, and the model can predict a time span T, where T is an integer greater than or equal to 0. Introducing the prediction fluctuation value γ, typically set as $\gamma = 1+5\%$, then taking the BI as an example, $X(t) \leq \gamma X_{balance1} \leq X(t+1)$. When the model parameters are determined not to meet the model conditions or when the regression analysis P-value is relatively large ($P>0.05$), dynamic prediction stops.

Modeling the full-cycle data of BI and VI allows for the generation of a complete forecast data table. Partial prediction results are shown in Fig. 2.

As shown in Fig. 2, applying the model to real Baidu Search Index and China's COVID-19 vaccination data validates the model's ability to predict future trends in online search behavior and actual vaccination behavior. In the figure, t represents the selected time interval length, and T represents the length of the future period that can be predicted. Two segments that can achieve the expected prediction targets well are identified: the first segment ranges from 192 days to 234 days, and the second segment ranges from 578 days to 639 days. In other words, using 25.3–36.6% of the entire cycle's data and 90.3–99.8% of the entire cycle's data yields better prediction results. For the prediction length T of BI, it falls between 92 days and 149 days, with an average prediction length of 106 days. Therefore, 25.3–36.6% of the BI for the entire cycle can provide relatively accurate predictions for 46–53% of the BI for the entire cycle. Similarly, for the prediction length T of VI, it falls between 38 days and 129 days, with an average prediction length of 55.77 days. Hence, 25.3–36.6% of the VI for the entire cycle can offer relatively accurate predictions for 38–45% of the VI for the entire cycle. In the second segment of forecasting length, both for BI and VI, the lengths of prediction are identical. Starting from the first 90.3% of the full-cycle data, the equation set fully covers all subsequent data for prediction. Additionally, both segments of prediction results exhibit significant regression analysis effectiveness, with the accuracy of some prediction outcomes shown in the Table 12.

Predicting public health events, such as the future trend of vaccine administration numbers, helps shorten the response time of public health professionals, enabling them to provide more effective services and thereby improving public health conditions. Traditionally, the prediction of public health events has primarily relied on clinical data, such as patient registration records. However, non-clinical network data, such as search engine queries, has been proven to be useful in predicting public health events. Clinical data and network data are complementary sources of evidence. While clinical data provide authoritative information recognized by experts, network data provide large-scale, near-real-time information, such as symptoms or health issues that may not have been discovered or reported through official clinical channels. This indicates that network data, such as search query frequencies, has potential utility in predicting vaccine administration situations in countries without national vaccine registration systems [33]. By

Table 11 Stability conditions for Equilibrium points

Equilibrium point	DET(J)	Tr(J)	Equilibrium condition
P0	$r_1 r_2$	r_1+r_2	Disequilibrium
P1	$DET(J(P1))$	$-(r_1+r_2)$	$\alpha+\beta > 1$
P2	$DET(J(P2))$	$-(r_1+r_2)$	$\alpha+\beta < 1$
P3	$-r_1 r_2$	$-(r_1+r_2)$	Disequilibrium
P4	$-r_1 r_2$	r_1-r_2	Disequilibrium
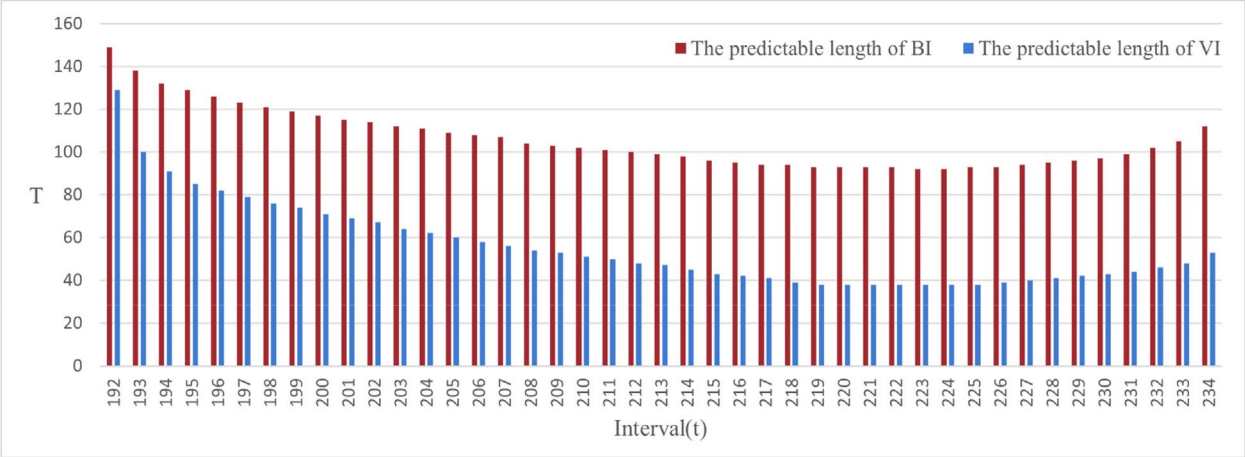
(a) the time interval ranges from 192 days to 234 days
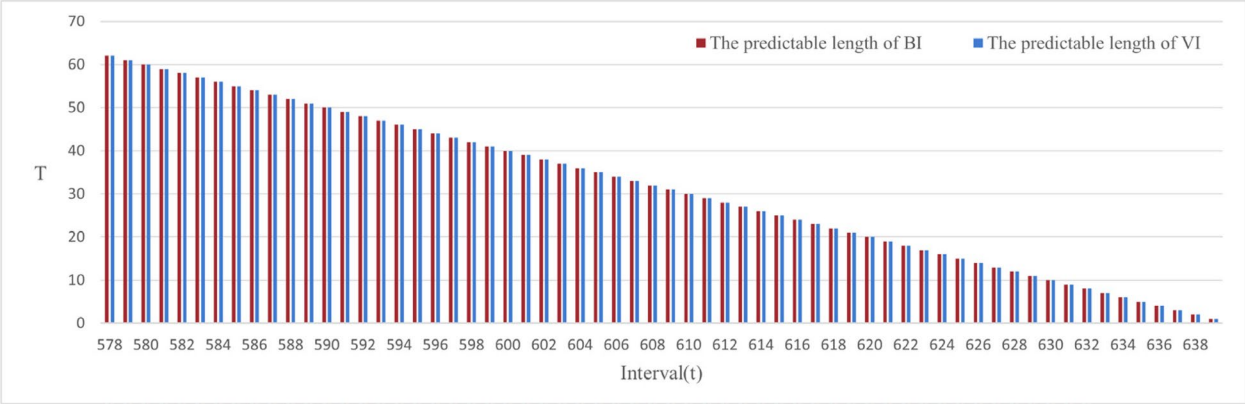
(b) the time interval ranges from 578 days to 639 days

Fig. 2 Partial prediction lengths for the entire data cycle

Table 12 The Accuracy of some prediction outcomes

Time	Prediction length T	*P*-value of regression analysis	R^2
2021.03.24–2021.10.01	149	$9.61 \times 10^{-36};8.51 \times 10^{-15};3 \times 10^{-4};1.34 \times 10^{-57};1.01 \times 10^{-26};1.03 \times 10^{-14}$	0.85; 0.94
2021.03.24–2021.10.02	138	$5.5 \times 10^{-36};6.27 \times 10^{-15};2 \times 10^{-4};2.91 \times 10^{-57};3.13 \times 10^{-26};3.24 \times 10^{-14}$	0.85; 0.94
2021.03.24–2021.10.03	132	$3.15 \times 10^{-36};4.57 \times 10^{-15};2 \times 10^{-4};6.03 \times 10^{-57};9.37 \times 10^{-26};9.86 \times 10^{-14}$	0.85; 0.94
2021.03.24–2022.10.22	62	$4.9 \times 10^{-156};6.9 \times 10^{-125};4.17 \times 10^{-98};1.51 \times 10^{-83};5.77 \times 10^{-36};3.69 \times 10^{-14}$	0.81; 0.72
2021.03.24–2022.10.23	61	$3.3 \times 10^{-156};5.1 \times 10^{-125};3.26 \times 10^{-98};1.09 \times 10^{-83};5.24 \times 10^{-36};3.8 \times 10^{-14}$	0.81; 0.94

continuously monitoring search trends, public health officials can identify areas with low vaccination rates and target interventions where they are most needed. This allows for the proactive allocation of resources to regions with increased vaccine hesitancy or misinformation, thus improving vaccination campaigns. In addition, real-time search trends can serve as early indicators of emerging health issues, allowing for swift policy responses to prevent the spread of misinformation or address new vaccine-related concerns. Moreover, the model's predictive

Liu *et al. BMC Public Health*   (2025) 25:975

Page 15 of 18

capabilities can support intervention design by offering insights into potential future vaccine uptake, enabling policymakers to craft tailored messaging, adjust resource distribution, and preemptively address public concerns. This integrated approach of monitoring, prediction, and intervention ensures a more dynamic and flexible response to public health crises, ultimately improving both the efficiency and effectiveness of public health interventions.

Conclusion

Building upon previous theoretical research, this paper divides the social system into online network data and offline social entities. In selecting research data that conforms to the associative impact patterns between network data and social entities in detecting social systems for quantitative analysis, considering that after certain sudden events occur, it may not be possible to quantify both network data and social entities simultaneously, or the impact of the events may be limited in time or space, rendering the quantitative data unrepresentative. Therefore, we ultimately chose the COVID-19 pandemic as the sudden public health event for analysis. This is because of its long time span, broad spatial scope, and high representativeness. This choice also aligns with the urgency of addressing public health challenges through real-time data analysis, where monitoring trends in internet searches could complement existing public health surveillance systems.

This study analyzed the associative impact patterns between internet searches and vaccine administration, constructing a detection model. Internet searches and vaccine administration were quantified as BI and VI, respectively. A logistic model was constructed to observe the fundamental evolutionary patterns. To measure the mutual influence between the two, an influence function was defined, and the best-fitting common structural factors were determined through data fitting, thereby constructing a nonlinear correlation mathematical model between the two.

Through research analysis, we have identified a symbiotic effect between internet searches and vaccine administration, uncovering a nonlinear correlation equation between the two. We have delved into the symbiotic effect between the two. For vaccine administration, extensive internet searches can enhance public understanding of relevant knowledge, positively influencing vaccination willingness. When monitoring changes in internet search trends, relevant authorities can gain timely insights into public attitudes and demands regarding vaccine administration, directly address these issues, and transform them into

insightful public health strategies, thereby aiding in the rational allocation of medical resources and strengthening public health management. This feedback mechanism provides a scientific basis for the formulation and adjustment of vaccine administration policies, helping to better meet public needs. For internet searches, monitoring changes in vaccine administration trends enables timely understanding of public needs and concerns regarding vaccination on internet search platforms, facilitating the provision of targeted health information related to vaccination, thereby enhancing the accuracy and timeliness of public access to health information. Therefore, national health system agencies promoting vaccination cannot avoid investing in internet dissemination, as it cannot be managed solely by private efforts but must be the result of the coordinated efforts of public health, private, and scientific associations, as well as social movements [34].

We applied a nonlinear correlation model between internet searches and COVID-19 vaccine to predictive research, indicating the method possesses a certain predictive capability. Currently, many countries still lack departments that promptly register vaccine administration information, relying on annual surveys for vaccine administration data collection. In these countries, estimating near real-time vaccine administration rates solely based on internet data is valuable. Vaccine administration rates can be automatically estimated from internet data rather than from slowly collected clinical records or population surveys [35]. Thus, this model contributes to predictive research on forecasting issues, aiding in formulating vaccine administration plans and providing predictive recommendations for dynamic adjustments in public health policies.

Limitations

(1) Data Sources: The dataset used in this study, sourced from Baidu, the predominant search engine in China with 648 million monthly active users as of December 2022, is highly representative of the Chinese population. The Baidu Index data collects from 34 provincial-level administrative regions provides a comprehensive overview of vaccine uptake across different regions, populations, and platforms in China. It accurately reflects the public healthcare demand of the Chinese population during the pandemic. However, this dataset cannot fully represent the vaccine uptake situation in other regions, especially in countries where different platforms or search engines are more commonly used.

(2) Time Frame: This study is based on all the COVID-19 vaccination statistics published by the National

Liu *et al. BMC Public Health* (2025) 25:975

Page 16 of 18

Health Commission of China, covering the entire period of the evolution of China's COVID-19 vaccination campaign, ensuring the representativeness of the analysis. However, challenges arise when modeling long-term behavioral shifts, especially as public attention towards COVID-19 and vaccination has diminished over time. As public concern wanes, the dynamics of search behaviors may change, which could affect the model's predictive accuracy in future scenarios. However, our team has noticed this situation and has already initiated further research [36].

(3) Search Behavior: This study addresses the nonlinear detection of the relationship between internet search trends and vaccine behaviors at the macro level, proposing a model for detecting nonlinear relationships between online and offline data. However, the paper lacks an in-depth analysis of individual search behaviors, such as search paths, stay time, and differentiation between positive and negative vaccine-related information. A more detailed examination of user behavior could provide deeper insights into how different types of content, particularly vaccine-related, influence public sentiment and behavior, thereby enhancing the predictive power of the model.

Research potentials

The method introduced in this study for detecting the interaction between online and offline data demonstrates significant universality and adaptability. This mathematical modeling approach is not restricted to a particular time frame or geographical context; rather, it provides a versatile framework applicable across diverse temporal, spatial, and data-related dimensions. The methodology can be extended to examine interactions in various research fields, such as public health, social sciences, and even entertainment. The model has already been successfully applied to a range of datasets, including COVID-19 case data coupled with Baidu search trends, as well as entertainment-related data such as movie box office statistics. These applications highlight the robustness and flexibility of the model, underscoring its potential for cross-domain analysis. Moving forward, the model can be applied to different regions and platforms, thus serving as a powerful tool for investigating a wide spectrum of socio-economic behaviors.

Furthermore, this study not only clarifies the mathematical mechanisms underlying the correlation between Internet search behaviors and vaccination rates but also presents an integrated model that combines both analytical and predictive functions. This dual-purpose nature makes the model an invaluable resource for policymakers

and researchers, as it offers both an analysis of current trends and forecasts of future developments. The predictive capability of the model enhances the agility of public health systems, particularly in facilitating real-time decision-making. By providing a deeper understanding of how online search behaviors mirror public health trends, the study sets the stage for further applications in forecasting health behaviors, vaccination trends, and even anticipating potential outbreaks of other diseases. This innovative approach contributes to the development of a new empirical and theoretical framework in public health research, providing fresh insights into how social phenomena can be modeled and predicted using readily available online data.

Future work

Building on the findings of this study, future research could focus on enhancing the model's accuracy and applicability. One direction is to introduce additional parameters, such as personalized search behaviors, including individual preferences, search frequency, and sentiment analysis of vaccine-related content (positive or negative). These factors could provide a deeper understanding of how online interactions influence public attitudes toward vaccination.

Expanding the model to include data from diverse search platforms, beyond Baidu, would offer a more comprehensive view of public health behaviors, particularly in regions where alternative search engines or social media platforms are prevalent. This cross-platform integration would strengthen the model's robustness and generalizability. Additionally, incorporating data from various online sources, such as forums, news outlets, and government portals, could yield more nuanced insights into public opinion and its relationship with health behaviors.

Further advancements could include developing an information index to quantify the impact of public sentiment on health behaviors, categorizing search trends as positive, negative, or neutral, and assessing their influence on vaccination actions. Understanding how these shifts in opinion occur over time could improve the model's predictive capabilities, enabling more responsive public health strategies.

Moreover, examining the temporal lag between online search trends and offline behaviors, such as vaccination rates, would enhance the model's predictive power. By integrating these advanced features, the model could serve as a valuable tool for real-time monitoring and decision-making, supporting not only vaccination campaigns but also broader health initiatives. This research would contribute to the growing body of work using digital data to inform public health policies, offering insights into the relationship between online behaviors and health outcomes.

Liu *et al. BMC Public Health*   (2025) 25:975

Page 17 of 18

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Shekhar R, Sheikh AB, Upadhyay S, Singh M, Kottewar S, Mir H, et al. COVID-19 vaccine acceptance among health care workers in the United States. Vaccines. 2021;9:119.
2. Marian AJ. Current state of vaccine development and targeted therapies for COVID-19: impact of basic science discoveries. Cardiovasc Pathol. 2021;50:107278.
3. Di Domenico G, Nunan D, Pitardi V. Marketplaces of misinformation: a study of how vaccine misinformation is legitimized on social media. J Public Policy Mark. 2022;41:319–35.
4. Nazli SB, Yigman F, Sevindik M, Ozturan DD. Psychological factors affecting COVID-19 vaccine hesitancy. Ir J Med Sci. 2022;191:71–80.
5. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457:1012–4.
6. Xie G, Li X, Qian Y, Wang S. Forecasting tourism demand with KPCA-based web search indexes. Tour Econ. 2021;27:721–43.
7. Yamaguchi S, Hinoki A, Tsubouchi K, Amano H, Tajima A, Uchida H. Usefulness of web search queries for early detection of diseases in infants. Nagoya J Med Sci. 2021;83:107–11.
8. Rizun N, Baj-Rogowska A. Can web search queries predict prices change on the real estate market? IEEE Access. 2021;9:70095–117.
9. Troelstra SA, Bosdriesz JR, de Boer MR, Kunst AE. Effect of tobacco control policies on information seeking for smoking cessation in the Netherlands: a Google trends study. PLoS One. 2016;11:e0148489.
10. Maugeri A, Barchitta M, Agodi A. Using Google trends to predict COVID-19 vaccinations and monitor search behaviours about vaccines: a retrospective analysis of Italian data. Vaccines (Basel). 2022;10:119.
11. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of Google trends in health care research: a systematic review. PLoS One. 2014;9:e109583.
12. Pullan S, Dey M. Vaccine hesitancy and anti-vaccination in the time of COVID-19: a Google trends analysis. Vaccine. 2021;39:1877–81.
13. Awijen H, Ben Zaied Y, Nguyen DK. Covid-19 vaccination, fear and anxiety: evidence from Google search trends. Soc Sci Med. 2022;297:114820.
14. Moussa OZ, Takeuchi K. Does searching online for vaccination information affect vaccination coverage? Evidence from sub-Saharan African countries. Econ Hum Biology. 2022;47:101181.
15. Wang S-C, Chen Y-C. Exploration of correlations between COVID-19 vaccination choice and public mental health using Google trend search. Vaccines. 2022;10:2173.
16. Díaz F, Henríquez PA, Hardy N, Ponce D. Population well-being and the COVID-19 vaccination program in Chile: evidence from Google trends. Public Health. 2023;219:22–30.
17. Lazebnik T, Bunimovich-Mendrazitsky S, Ashkenazi S, Levner E, Benis A. Early detection and control of the next epidemic wave using health communications: development of an artificial intelligence-based tool and its validation on COVID-19 data from the US. Int J Environ Res Public Health. 2022;19:16023.
18. Lazebnik T. Computational applications of extended SIR models: a review focused on airborne pandemics. Ecol Model. 2023;483:110422.
19. Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Gavrilov D, Giattino C, Joe Hasell J, Macdonald B, Dattani S, Beltekian D, Ortiz-Ospina E, Roser M. Coronavirus (COVID-19) Vaccinations. 2020. Published online at OurWorldinData.org. Retrieved from: https://ourworldindata.org/covid-vaccinations. Online Resource.
20. Middleman AB, Klein J, Quinn J. Vaccine hesitancy in the time of COVID-19: attitudes and intentions of teens and parents regarding the COVID-19 vaccine. Vaccines (Basel). 2021;10:4.
21. Dinleyici EC, Borrow R, Safadi MAP, van Damme P, Munoz FM. Vaccines and routine immunization strategies during the COVID-19 pandemic. Hum Vaccin Immunother. 2021;17:400–7.
22. Robinson R, Nguyen E, Wright M, Holmes J, Oliphant C, Cleveland K, et al. Factors contributing to vaccine hesitancy and reduced vaccine confidence in rural underserved populations. Humanit Soc Sci Commun. 2022;9:1–8.
23. de Albuquerque Veloso Machado M, Roberts B, Wong BLH, van Kessel R, Mossialos E. The relationship between the COVID-19 pandemic and vaccine hesitancy: a scoping review of literature until August 2021. Front Public Health. 2021;9:747787.
24. Colautti L, Cancer A, Magenes S, Antonietti A, Iannello P. Risk-perception change associated with COVID-19 vaccine's side effects: the role of individual differences. Int J Environ Res Public Health. 2022;19:1189.
25. Bik HM, Goldstein MC. An introduction to social media for scientists. PLoS Biol. 2013;11:e1001535.
26. Fobiwe JP, Martus P, Poole BD, Jensen JL, Joos S. Influences on attitudes regarding COVID-19 vaccination in Germany. Vaccines (Basel). 2022;10:658.
27. Paterson P, Meurice F, Stanberry LR, Glismann S, Rosenthal SL, Larson HJ. Vaccine hesitancy and healthcare providers. Vaccine. 2016;34:6700–6.
28. Bish A, Yardley L, Nicoll A, Michie S. Factors associated with uptake of vaccination against pandemic influenza: a systematic review. Vaccine. 2011;29:6472–84.
29. Signorelli C, Odone A. Advocacy communication, vaccines and the role of scientific societies. Ann Ig. 2015;27:737–47.
30. Thielmann A, Viehmann A, Weltermann BM. Effectiveness of a web-based education program to improve vaccine storage conditions in primary care (keep cool): study protocol for a randomized controlled trial. Trials. 2015;16:301.
31. Bragazzi NL, Barberis I, Rosselli R, Gianfredi V, Nucci D, Moretti M, et al. How often people google for vaccination: qualitative and quantitative insights from a systematic search of the web-based activities using Google trends. Hum Vaccines Immunotherapeutics. 2017;13:464–9.
32. Xia Y, Li Q, Jiao W, Lan Y. Dynamic mechanism of eliminating COVID-19 vaccine hesitancy through web search. Front Public Health. 2023;11:1018378.
33. Dalum Hansen N, Lioma C, Mølbak K. Ensemble learned vaccination uptake prediction using web search queries. In: Proceedings of the 25th ACM

Liu *et al. BMC Public Health*   (2025) 25:975

Page 18 of 18

International on Conference on Information and Knowledge Management. New York: Association for Computing Machinery; 2016. p. 1953–6.

34. Tafuri S, Gallone MS, Gallone MF, Zorico I, Aiello V, Germinario C. Communication about vaccinations in Italian websites: a quantitative analysis. Hum Vaccines Immunotherapeutics. 2014;10:1416–20.

35. Dalum Hansen N, Mølbak K, Cox IJ, Lioma C. Time-series adaptive estimation of vaccination uptake using web search queries. In: Proceedings of the 26th International Conference on World Wide Web Companion. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee; 2017. p. 773–4.

36. Wang Y, Ran L, Jiao W, Xia Y, Lan Y. The predation relationship between online medical search and online medical consultation—empirical research based on Baidu platform data. Front Public Health. 2024;12:1392743.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.