

# Social inequality and cultural factors impact the awareness and reaction during the cryptic transmission period of pandemic

Zhuoren Jiang<sup>a</sup>, Xiaozhong Liu<sup>b,\*</sup>, Yangyang Kang<sup>c,d,e</sup>, Changlong Sun<sup>c,d</sup>, Yong-Yeol Ahn<sup>f</sup> and Johan Bollen<sup>f</sup>

<sup>a</sup>School of Public Affairs, Zhejiang University, 866 Yuhangtang Rd, Hangzhou, Zhejiang 310058, P.R. China

<sup>b</sup>Data Science, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA

<sup>c</sup>Tongyi Lab, Alibaba Group, 969 West Wen Yi Road, Hangzhou, Zhejiang 310030, P.R. China

<sup>d</sup>College of Computer Science and Technology, Zhejiang University, Yuquan Campus, Hangzhou, Zhejiang 310013, P.R. China

<sup>e</sup>Polytechnic Institute, Zhejiang University, 866 Yuhangtang Rd, Hangzhou, Zhejiang 310058, P.R. China

<sup>f</sup>Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, 107 S. Indiana Avenue, Bloomington, IN 47405, USA

\*To whom correspondence should be addressed: Email: [xliu14@wpi.edu](mailto:xliu14@wpi.edu)

Edited By Rebecca Seguin-Fowler

## Abstract

The World Health Organization (WHO) declared the COVID-19 outbreak a Public Health Emergency of International Concern on 2020 January 31. However, rumors of a “mysterious virus” had already been circulating in China in 2019 December, possibly preceding the first confirmed COVID-19 case. Understanding how awareness about an emerging pandemic spreads through society is vital not only for enhancing disease surveillance, but also for mitigating demand shocks and social inequities, such as shortages of personal protective equipment and essential supplies. Here we leverage a massive e-commerce dataset comprising 150 billion online queries and purchase records from 94 million people to detect the traces of early awareness and public response during the cryptic transmission period of COVID-19. Our analysis focuses on identifying information gaps across different demographic cohorts, revealing significant social inequities and the role of cultural factors in shaping awareness diffusion and response behaviors. By modeling awareness diffusion in heterogeneous social networks and analyzing online shopping behavior, we uncover the evolving characteristics of vulnerable populations. Our findings expand the theoretical understanding of awareness spread and social inequality in the early stages of a pandemic, highlighting the critical importance of e-commerce data and social network data in effectively and timely addressing future pandemic challenges. We also provide actionable recommendations to better manage and mitigate dynamic social inequalities in public health crises.

**Keywords:** pandemic awareness modeling, pandemic preparedness, awareness inequality, eCommerce data analytics, heterogeneous social networks

## Significance Statement

This study comprehensively characterizes how emerging disease information spread through the Chinese population by examining large-scale chronological purchasing data of personal protective equipment on an eCommerce platform. Analyzing data from 94 million individuals, it reveals how socioeconomic status, social networks, geography, and cultural factors relate to awareness diffusion, uncovering substantial inequalities in information access and response. These findings expand the theoretical framework on awareness diffusion and social inequality during pandemics and highlight evolving characteristics of vulnerable populations, emphasizing early interventions tailored to specific communities. The study also provides practical recommendations for adopting flexible and dynamic public health strategies and offers valuable insights into the complex influence of cultural factors and social structures on pandemic responses.

## Introduction

The emergence of COVID-19 was accompanied by demand surges for PPEs (Personal Protective Equipments) and other consumer goods in nearly every country (1). For instance, China experienced shortages of sanitizing wipes, face-masks, and other essential goods in early 2020 (2, 3). Although official confirmation of the disease drastically accelerated demand, this demand surge began

well before the official WHO announcement of a Public Health Emergency of International Concern (PHEIC), due to rumors of a “mysterious virus” on social media that started circulating as early as December 2019 (4, 5). This suggests that the information about an emerging pandemic can spread well before its official confirmation during its cryptic transmission period, and that the traces of such information may be detectable in rapid changes

**Competing Interest:** The authors declare no competing interests.

**Received:** February 5, 2024. **Accepted:** January 3, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

in demand for particular consumer goods. Although it has been several years since COVID-19 has become a pandemic, the relevance of examining this phenomenon persists, as we anticipate the next pandemic and socioeconomic inequalities continue to shape how different demographics are affected by such shifts in demand and consumption, with vulnerabilities dynamically evolving based on the spread of awareness and information throughout the course of a pandemic. Hence, we examine behavioral data captured by a massive e-commerce platform during 88 days from 2019 December 1 to 2020 February 26, which is referred to as the “cryptic transmission period.”<sup>a</sup>

To understand this phenomenon tying early awareness of a pandemic during its cryptic period to changes in consumer behavior and concomitant social inequities, we need to address two questions: (i) “how does the awareness of an emerging pandemic spread through the various demographics of a population?” and (ii) “how does authoritative and officially vetted information from local, national, and international instances affect awareness diffusion?” Understanding the dynamics of pandemic awareness during the cryptic period can critically inform emerging disease surveillance, but it may also shed light on how demand surges and shortages of critical goods affect the public response to future health emergencies. It is particularly important to understand how the differential diffusion of awareness creates information gaps between different demographic cohorts, which can lead to socioeconomic inequities that negatively affect a society’s ability to respond to an emerging public health crisis involving an infectious disease (6–8). In fact, in the case of the COVID-19 pandemic, shortages of PPEs and essential supplies, or the lack of equitable distribution, may have disproportionately affected disadvantaged population groups potentially contributing to higher infection rates and mortality (9, 10). In a public health emergency, enhancing public awareness and preparedness, and doing so in a socially and economically equitable manner, can be as important as mitigating the crisis itself (11, 12).

Here, we investigate the early spread of pandemic awareness in China by data mining a unique record of 150 billion query and purchase records for 94 million individuals from the largest eCommerce platform, Alibaba. These data were collected right at the crucial time period of the early COVID-19 outbreak (2019 December 1 to 2020 February 26), along with longitudinal records of purchasing history of the included individuals in the past 10 years, providing detailed geo-located shipping/gifting information from which 43.7 million families, 120.3 million schoolmates, and 25.8 million workmates relations were informed (13). By focusing on PPE-related queries and orders, these data enable investigations into the collective querying and purchasing behavior at the cryptic periods of the pandemic, leveraging changes in consumer demand as a proxy of awareness (11, 14). We examine the dynamic factors that are associated with the timing of the initial response for a variety of vulnerable populations in terms of their location, social relations, and demographics.

The eCommerce data recorded in this period provide a unique opportunity to study the evolving pandemic awareness throughout various socioeconomic demographics for the following reasons. First, eCommerce is now the primary means of shopping for the Chinese population.<sup>b</sup> Second, eCommerce data collection reflects user behavior in its natural state. This nonintrusive approach offers an unbiased view of an individual’s actual needs and preferences, mitigating social conformity and observer effects typically associated with traditional self-reported data. Third, unlike social media or search engine data, which often reflect only interactions and textual expressions, eCommerce data capture a substantial, real-

time, and representative snapshot of actual purchasing behaviors, geographic distribution, and consumption patterns. It can enable comprehensive investigations into the associations between individual demographics and purchasing behaviors, which can be further enhanced by detecting longitudinal patterns. For instance, if an individual purchased pregnancy products 7 years ago, diapers 6 years ago, and toys recently, it can be inferred that the individual may have a child of age 7. This helps us analyze the connections between demographic factors, pandemic awareness, and public health risks. This allows the estimation of rich information to unravel associations between different demographic factors, pandemic awareness, and a variety of public health risks. Fourth, purchase records also contain gifting and shipping information, which enables multiview social ties estimation. For example, when two people share the same home shipping address, we may infer a social or familial bond (13). Furthermore, when shipping addresses pertain to the same company/school/dorm address, we can infer individuals’ classmate/workmate relations (13). This enables a detailed analysis of social networks and their impact on pandemic awareness and disaster planning. By leveraging these unique data characteristics, eCommerce data facilitate innovative research that can track the diffusion of pandemic awareness with high resolution across social, temporal, demographic, and geographical dimensions, at truly societal scales (a majority of the Chinese population).<sup>c</sup> These data also allow for early detection of shifts in consumer behavior related to health concerns, providing valuable signals for timely public health interventions.

## Data collection and awareness label generation

We analyze how early awareness of COVID-19 spread through the various sectors, locations, social networks, and demographics of the Chinese population by leveraging a data set of 46.5 billion queries (randomly sampled) issued by 800 million individuals, covering 88 days from 2019 December 1 to 2020 February 26, a crucial period around the time WHO declared a public health emergency, here referred to as the “cryptic transmission period.” In addition, for the same population, we randomly sampled 150 billion historical queries and purchase behaviors, shipping and gifting addresses, etc. ranging from 2010 to 2019, to estimate individuals’ demographic variables and their social relationships, i.e. family, workmate, and schoolmate networks (see [Supplementary Material, 2.C](#)).

Our reasons for selecting this time period and China as the focus of our research are as follows: First, China is the first country to report a large-scale outbreak of COVID-19 (15), with the initial confirmed cases reported in December 2019 in Wuhan, Hubei Province (16, 17). As the first country to confront the pandemic, China’s initial response provides invaluable insights into societal awareness and reactions during the crucial early stages of a global health crisis. Second, the chosen time period reflects distinctive social behavior responses. During this period, public knowledge of the virus was limited (17, 18), and guidance from the government and healthcare institutions gradually became clearer (15). Concurrently, the crucial resources were potentially being rapidly and unequally consumed (10, 19). Investigating public awareness during “cryptic transmission period,” characterized by incomplete information and high uncertainty, will help effectively identify vulnerable populations and understand how to enhance societal resilience in future crises, mitigating the potentially severe consequences of inequality. Third, China’s vast and complex social structure, characterized by wealth disparities, high population

density, and significant population mobility, presents unique challenges. Studying how Chinese communities and individuals became aware and responded in this complex environment can provide valuable lessons for crisis management in similarly complex social settings. In conclusion, China's "cryptic transmission period" during the COVID-19 pandemic offers a unique national context that can provide invaluable insights for global public health management and crisis response strategies.

We mark each individual  $i$  at time  $t$  as either "aware" ( $L_i^t = 1$ ) or "not aware" ( $L_i^t = 0$ ) following the definition of "pandemic awareness" proposed in (11).  $L_i^t$  is determined on the basis of whether they submitted any awareness-indicating query ( $q^* \in \mathbb{Q}$ ) three or more times at time  $t$ . We define a set of awareness-indicating queries  $\mathbb{Q}$  that are deemed related to COVID-19 PPEs, i.e. "(n95 or kn95 or kf94) & Face mask." All queries in  $\mathbb{Q}$  were found to discriminate well between pandemic-related and nonpandemic-related search activity; the queries were nearly never issued before the start of the pandemic and experienced dramatic growth during the early pandemic, with negligible noise and an average increase (in search population) of 102,792% compared with the same period from 2018 to 2019. Once  $L_i^t = 1$ , individual's awareness status will not revert. Only individuals with active eCommerce behaviors (who made at least a purchase every month in the past 5 years) were employed for awareness modeling. Following these filtering guidelines, we located 94,534,663 qualified individuals (11.8% of all the 800 million individuals) for awareness analysis and model generation. By 2020 February 26, the last day of our observation period, all individuals' awareness labels had changed to 1. This study was reviewed by the IRB at Indiana University (Protocol No. 10521) and was determined not to constitute human subjects research, thus not requiring further review. The experiment and secured data processing methods were also reviewed and approved by the Alibaba legal department. All data used in this study have been anonymized and de-identified to prevent any possibility of identifying individuals.

To provide exhaustive awareness analysis, following WHO guidance (20), we segment our COVID-19 cryptic transmission period into five subphases,<sup>d</sup> as shown in Table 1.

## Societal spread of COVID-19 awareness

Figure 1a illustrates daily (dotted line) and cumulative (dash-dotted line) growth of the number of individuals whose status changed to "aware." The trend exhibits rapid growth between January 19 and 26 (after "National Health Commission of China confirmed human-to-human transmission"), and experiences two peaks on January 23 ("Wuhan lockdown") and January 25 ("China activated first-level public health emergency"). Note that since January 27, most PPEs (e.g. face masks) were out of stock in China, implying that individuals who changed status to "aware" afterwards might not have been able to obtain PPE.

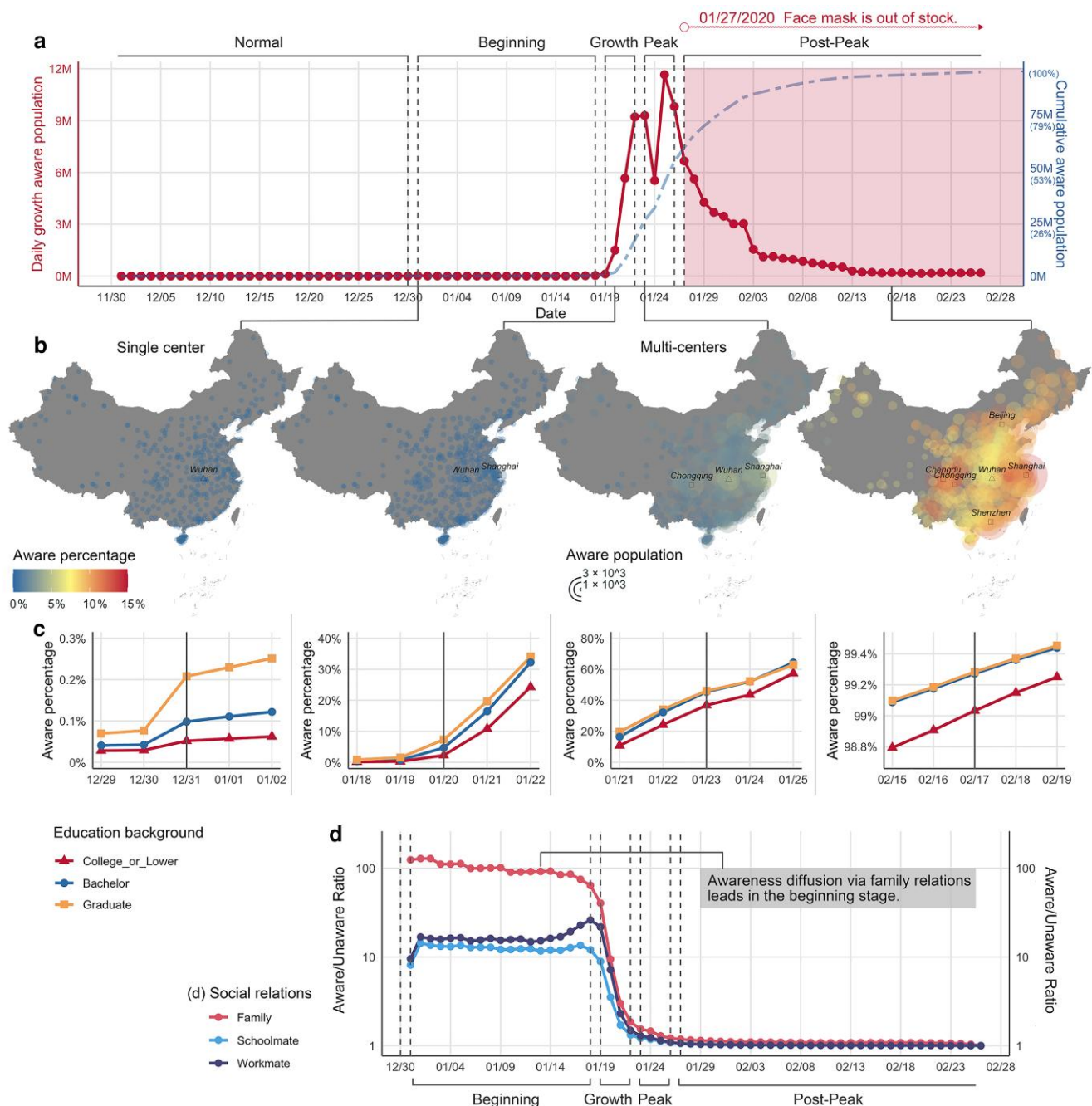
## Geography, education, and social relation analysis

Figure 1b visualizes the geographic distribution of the aware population and the awareness percentages of the populations of 366 major cities in Mainland China over time. During the normal and beginning phases (see Table 1), the epicenter Wuhan had the most aware individuals. The number of aware individuals in Shanghai (a leading metropolis) increased significantly in the growth phase. At the peak and post-peak phases, the awareness percentage also increased rapidly in other major cities. Geographically, awareness was initially correlated with the severity of the pandemic, but subsequently spread across the entire country, while the population's response seemed to be most pronounced in large cities.

Figure 1c shows awareness levels across educational backgrounds. In the beginning phase, the postgraduate group (master or PhD) exhibited the highest awareness percentage which continued into the growth phase, revealing that educational background could be an important variable driving early pandemic awareness or preparedness. Highly educated demographics seem to have had an advantage in picking up on news and rumors of the epidemic effectively, resulting in earlier and therefore more extensive access to scarce resources when they were still available. Throughout the peak and post-peak phases, individuals with postgraduate and bachelor degrees share similar awareness patterns. Individuals with lower educational levels exhibit the lowest awareness

**Table 1.** Five different phases (20) of the cryptic period of COVID-19 with specific real-world events

Phase	Time period	Definition	News/Social events
Normal	2019 December 1 to 2019 December 30	All province-level awareness percentages $\leq 0.001\%$	2019 December 8: A COVID-19 case was reported in retrospective studies (17)
Beginning	2019 December 31 to 2020 January 18	First province-level awareness growth rate $> 100\%$ and national awareness percentage $> 0.001\%$	2019 December 31: Wuhan MHC released a briefing (pneumonia outbreak) 2020 January 5: Wuhan MHC reported 59 cases of viral pneumonia 2020 January 16: Strict exit screening measures activated in Wuhan
Growth	2020 January 19 to 2020 January 22	Second province-level awareness growth rate $> 100\%$ and national awareness percentage $> 0.001\%$	2020 January 20: China NHC confirmed human-to-human transmission
Peak	2020 January 23 to 2020 January 26	More than 95% provinces' awareness growth rates $> 10\%$ and national awareness percentage $> 0.1\%$	2020 January 23: Wuhan lockdown 2020 January 24: Hubei activated first-level public health emergency 2020 January 25: China activated first-level public health emergency
Post-Peak	2020 January 27 to 2020 February 26	More than 95% provinces' awareness growth rates $< 10\%$ and continuously drop for at least 3 days	2020 January 31: WHO declared the novel coronavirus outbreak a PHEIC 2020 February 2: Wuhan launched quarantine strategies 2020 February 11: WHO named COVID-19 officially



**Fig. 1.** The patterns of awareness diffusion (five Phases). **a**) The diffusion of awareness and reaction. The Y-axis on the left represents the daily growth in the aware population (dotted line), and the Y-axis on the right corresponds to the daily cumulative aware populations (dash-dotted line). There are two peaks on the daily trends: 9,289,545 newly aware on 2020 January 23 (Wuhan lockdown) and 11,655,320 newly aware on 2020 January 25 (30 provincial-level regions activated first-level public health emergency). **b**) The geographic awareness distributions (366 cities) on four representative days of different phases. The size of the circle indicates aware population, the color indicates the awareness percentage, the triangle represents the epicenter, and the squares represent the city with most aware individuals. The initial awareness surged from the epicenter Wuhan (in the beginning phase), and gradually spread across the whole country with the increasing in pandemic severity. **c**) The awareness percentage trends of different education groups (for four representative days of different phases). In the beginning phase, the group with graduate degrees led a higher aware percentage after Wuhan MHC released a pneumonia outbreak briefing (2019 December 31). The similar trends can be observed in growth phase when China NHC confirmed human-to-human transmission (2020 January 20). During the peak phase (after Wuhan lockdown), the aware percentages of graduate and bachelor groups were close and higher than the college or lower group. This trend continued during the post-peak phase. **d**) Neighborhood awareness ratio (between aware individuals' aware neighbor percentage and unaware individuals' aware neighbor percentage) following three types of social relations. In the beginning phase, all three social relations can diffuse pandemic information efficiently (ratios  $>8$ ), while family relation showed the highest diffusion efficiency (ratios  $\in [63.9, 128.7]$ ). After the growth phase, it is hard to differentiate aware and unaware individuals' neighbors (ratio converges to 1). The basemap used in Figure 1 is sourced from the National Platform for Common Geospatial Information Service, China (Map Approval Number: GS (2024) 0650) and complies with Chinese laws and regulations. This map is for illustrative purposes only, used to visualize research data without any political or territorial assertions.



percentage across all phases (e.g. on average, the awareness percentage of the postgraduate group is 4.39 times that of the college or lower group during the beginning phase). As a result, they are likely to have responded more slowly to the emerging pandemic, and may have been more vulnerable in terms of pandemic preparedness and access to PPE.

In Fig. 1d, we explore changes in how awareness propagates through different types of social relations, by defining a “social neighborhood awareness ratio,” i.e. the ratio of the aware neighborhood percentage of aware individuals to that of unaware individuals across different phases (see [Supplementary Material, 3.D](#) for the detailed math definition). In this context, the term “social neighbors” refers to individuals who are directly connected to a person through a particular social relationship. We infer familial, work, and educational relations (represented by the *family*, *workmate*, and *schoolmate* networks, respectively) in this study. For each social relation type, the neighborhood awareness ratio reveals the degree to which awareness diffuses through the specific type of relation. From Fig. 1d, we observe that, in the beginning phase, familial relations can be particularly influential for information diffusion, and work relations can be more important than educational relations. In the later stages, the potential impact of all these relations may have declined rapidly due to the aggressive speed with which the pandemic spread across society.

Changes in awareness across the different social relationship types can be informative. Note that one cannot use these data to infer causality, due to the (latent) homophily (21). We caution against overinterpreting these differences, as each type of social tie may be influenced by varying levels of homophily and other confounding variables.

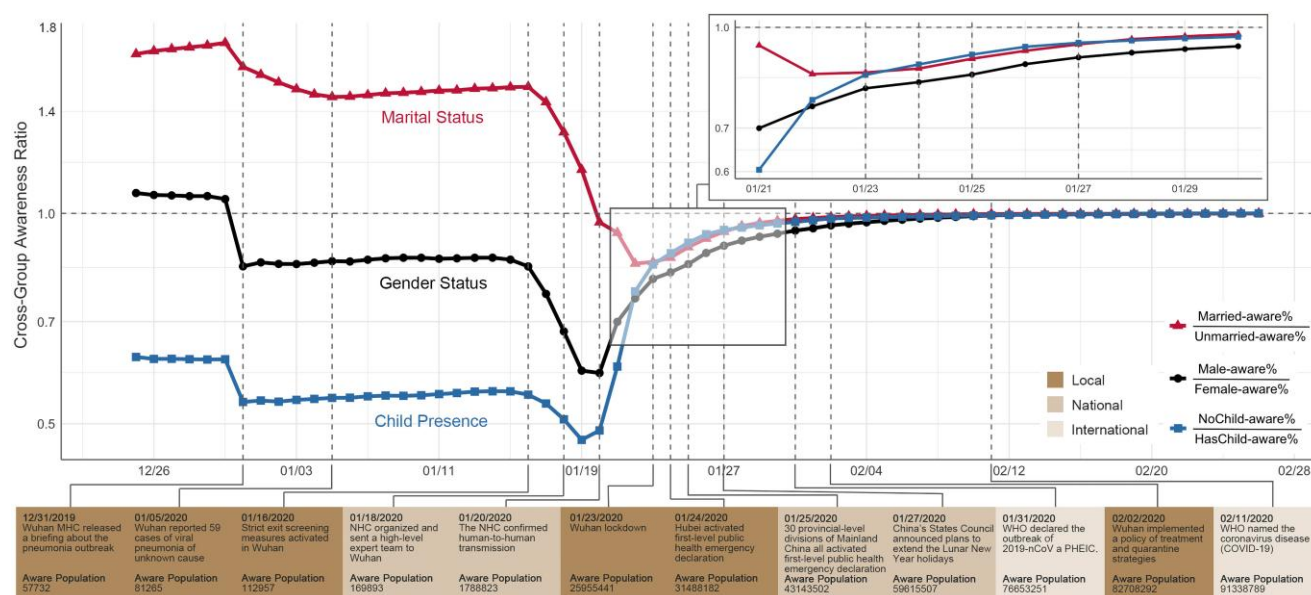
## Gender, marital, and child-presence analysis

Figure 2 shows differences between early awareness changes according to marital, gender, and child-presence status. The cut-out

zooms in on details for 10 days after the NHC (National Health Commission of China) confirmed human-to-human transmission. The Male-aware/Female-aware ratio trend line shows how awareness changed over time according to gender. Notably, the ratio does not change monotonically, indicating that certain events can trigger differential effects in male and female populations. For instance, the two events when (i) “Wuhan MHC (Wuhan Municipal Health Commission) released a briefing about the pneumonia outbreak” (2019 December 31) and (ii) when “strict exit screening measures were activated in Wuhan” (2020 January 16) corresponded to a drop in the Male-aware/Female-aware ratio (from 1.05 to 0.6), whereas the event of “The NHC confirmed human-to-human transmission” (2020 January 20) corresponded to a sharp increase in the ratio. This phenomenon suggests that women may be more sensitive or receptive to indications of an emerging pandemic. This gender gap however seems to be narrowing as the severity of the pandemic and public awareness increases, echoing prior studies (22–24) which show gender differences in risk perception and risk aversion.

NoChild-aware/HasChild-aware ratio trajectory shows that individuals with children are much more likely to react to the emerging pandemic than those without children. This is observed throughout the entire data coverage (ratio  $\leq 1$ ), with the ratio reaching 0.48 before “The NHC confirmed human-to-human transmission.” This ratio trend dropped 13% when “Wuhan MHC released a briefing about the pneumonia outbreak” (2019 December 31), indicating that individuals with children have greater awareness of an emerging pandemic during the beginning phase of cryptic transmission period.

The trend of Married-aware/Unmarried-aware ratio reveals that a higher percentage of individuals in the married group would search PPEs, well before “Wuhan MHC released a briefing on the pneumonia outbreak in the city” (peak on December 30, with ratio = 1.73). As pandemic concerns diffused through the



**Fig. 2.** The awareness trends across different demographic groups with important news events. Each line represents cross-group awareness ratio ( $R = P_{G_1}/P_{G_2}$ .  $R$  is the cross-group awareness ratio, and  $P_{G_i}$  is the percentage of aware people in the group  $G_i$ ). When the first official pandemic briefing released (2019 December 31), females, with-children, and unmarried groups reacted more quickly (cross-group awareness ratio trend-lines dropped). After strict screening tests were activated in Wuhan (2020 January 16), females, with-children, and unmarried groups showed stronger awareness strengths (awareness ratio trend-lines dropped and kept declining). After the NHC confirmed human-to-human transmission (2020 January 20), male and without-children groups began to show significant awareness strengths (awareness ratio trend-lines began to rise). It was not until the Wuhan lockdown (2020 January 23) that the married group began to show a relatively stronger level of awareness compared with the unmarried group (married-aware/unmarried-aware ratio trend-line began to rise). A base-10 log scale is applied for the Y-axis.

entire population, the growth rate of awareness in the unmarried group accelerated. However, it was not until “Wuhan lockdown” (2020 January 23) that the ratio began to rise again, indicating a resurgence in the relative awareness growth of the married group compared with the unmarried group.

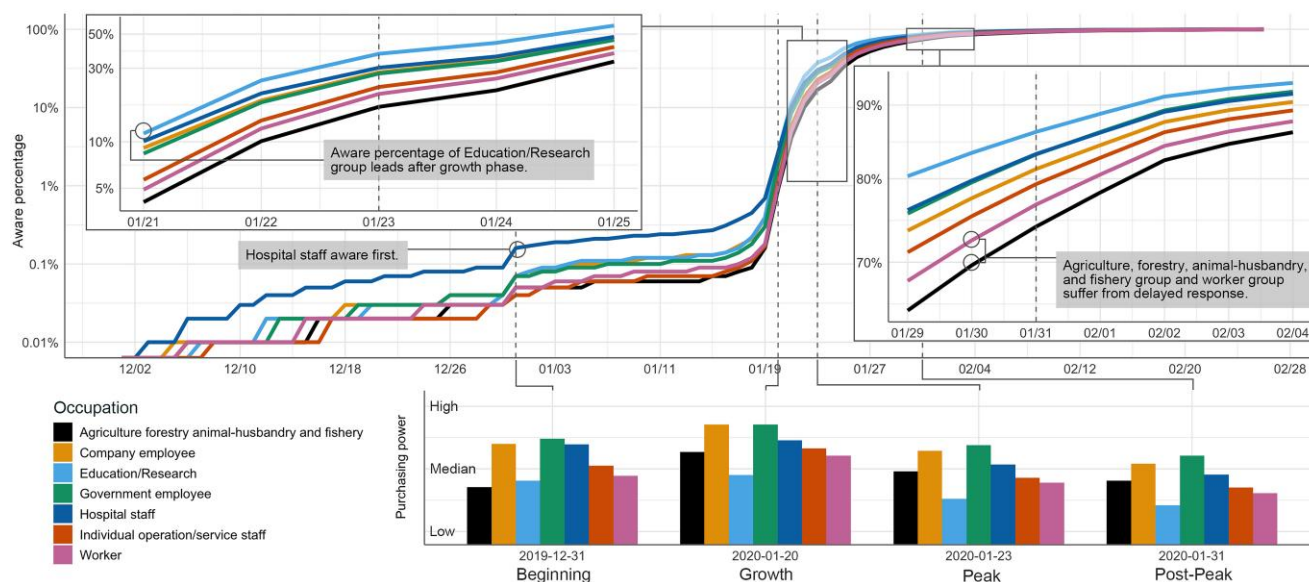
## Occupation and purchasing power analysis

Occupation and purchasing power (indicating income) inferred from eCommerce data reveal economic and social inequality during pandemic early stages.

As shown in Fig. 3, the awareness percentage of hospital staff leads other occupations during the normal and beginning phases (0.16–0.45%). When “Wuhan MHC released a briefing on the pneumonia outbreak” (on 2019 December 31), hospital staff became aware more rapidly compared with other communities. The awareness percentage of the research and education group which includes teachers, researchers, and students (2.67%) surpassed that of hospital staff group (2.65%) on 2020 January 20, when “China NHC confirmed human-to-human transmission.” Throughout the growth phase (2020 January 19 to 2020 January 22), the research and education group (0.39–25%), hospital staff (0.69–20.6%), white-collar employees (0.39–18.5%), and government employees (0.3–18.02%) maintained a high awareness percentage, whereas the awareness of “agriculture, forestry, animal husbandry, and fishery” group (0.16–10.9%), blue-collar workers (0.18–12.19%), and individual operation and service staff (0.17–13.74%) grew more slowly. The peak phase (2020 January 23 to 2020 January 26) shared a similar pattern on 2020 January 26, after “all 30 provincial-level regions activated first-level public health emergency.” The awareness percentage of research and education (66.25%), hospital staff (58.53%), government employees (56.89%), white-collar employees (56.13%), and individual operation and

service staff (52.38%) exceeded 50%, but the awareness percentages of agriculture, forestry, animal husbandry and fishery (44.03%) and blue-collar workers (48.38%) grew more slowly. It was not until “WHO officially named COVID-19” on 2020 February 11 that the awareness percentages of all occupation groups exceeded 95%. We also estimated individual purchasing power as a proxy to their income levels, enabling a comparison of awareness levels across income levels, which is crucial to gauge socioeconomic inequities. For all occupations, high-income groups seemed to have become aware of pandemic more quickly than low-income ones. We visualize the purchasing power of aware individuals on four representative days from each phase in Fig. 3 which highlights the income gap between communities that became aware earlier and later. Based on daily awareness growth rates, we found that the highest purchasing power groups (levels 6 and 7) maintained high growth across the beginning, growth, and peak phases. In contrast, the lowest purchasing power group (level 1) was slower to be aware, only becoming the fastest-growing in the later stages (after 2020 January 25). Among occupations, “hospital staff” and “education/research” showed rapid awareness growth in the earlier phases; while “workers” and “agriculture, forestry, animal husbandry, and fishery” were slower to be aware, only becoming the fastest-growing groups after 2020 January 27. When comparing the impact of occupation and purchasing power on awareness, occupational groups generally showed higher daily maximum growth rates, leading on 64 out of 88 days. This suggests that occupation may have a stronger influence on awareness than purchasing power. Detailed information on awareness percentages and daily awareness growth rates is available in [Supplementary Material, 2.D](#).

Overall, in the beginning stage of the pandemic, we found that high-income, well-educated individuals, and females became aware earlier than other groups, potentially affording them better



**Fig. 3.** Awareness (percentage) patterns for different occupation groups (upper subplot); a base-10 log scale is applied for the Y-axis; the left cut-out zooms in on details for 5 days around the Wuhan lockdown (2020 January 23); the right cut-out zooms in on details for 7 days around WHO declared the new coronavirus outbreak (2020 January 31) in the post-peak phase. The hospital staff kept the highest awareness percentage (0.16–0.45%) in the whole beginning phase. In the growth phase, the education/research group surpassed hospital staff and became the most aware group (0.39–25%), while agriculture forestry animal-husbandry and fishery were the least group (0.16–10.09%). The peak phase showed a similar pattern; education/research was the most aware group (37.24–66.25%) while agriculture forestry animal-husbandry and fishery were the least one (16.86–44.03%). The gaps of different occupation groups shrank during the post-peak phase. The lower subplot visualizes four representative days of different phases, and the Y-axis is the average purchasing power of the aware population from different occupation groups. Results show that high-income people respond to the emerging pandemic more quickly than low-income people.

opportunities to access and purchase scarce resources. Meanwhile, social relationships, familial in particular, may have played a key role in the diffusion of awareness.

## Geographic socioeconomic, cultural, and structural factor analysis

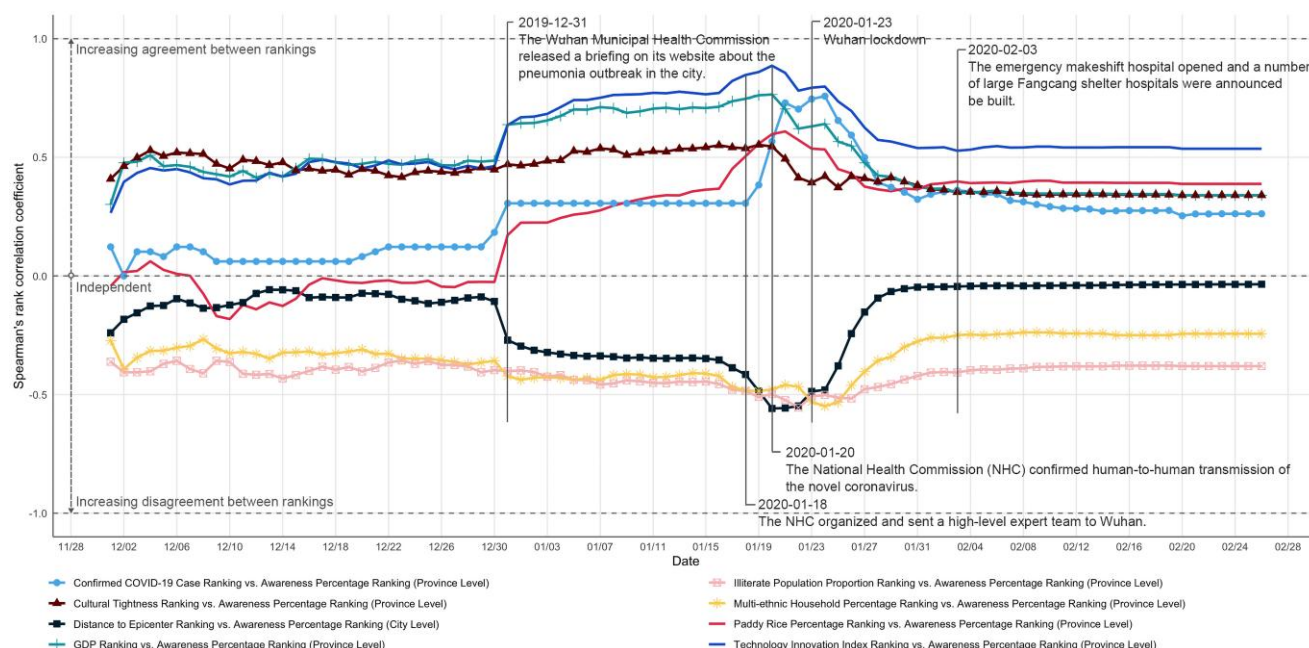
In terms of the geographic distribution of awareness dynamics (Fig. 1b), we found that awareness is strongly related to the geographic location. This observation inspired us to further explore the geographic factor(s) that shaped pandemic awareness, and their dynamics. We first focus on three main geographical factors: distance to the pandemic epicenter (Wuhan), confirmed COVID-19 case numbers, and local Gross Domestic Product (GDP). We rank the geographic locations based on the mentioned three factors and compare these to their ranking based on population awareness percentages for each day in the time period under consideration. See [Supplementary Material, 4.C](#) for detailed information.

We measured the statistical dependence between factor-based geo-rankings and awareness percentage geo-rankings by calculating their Spearman rank-order correlation (25). Figure 4 visualizes the trends of these three rank correlation coefficients. Overall, a city's awareness percentage and its distance to the epicenter are negatively correlated, and a province's confirmed COVID-19 cases and GDP (province rankings) are positively correlated with awareness. The degree of positive correlation between cases and awareness began to increase on 2019 December 31 (0.31) and rose rapidly from 2020 January 19 (0.38), 1 day after "the NHC organized and sent a high-level expert team to Wuhan." It peaked on 2020 January 24 (0.76), when "Hubei activated the first-level public health emergency," one day after the "Wuhan lockdown." Subsequently, the coefficient decreased from 0.76 to 0.26. The correlation coefficient between GDP and awareness percentage

exhibited a similar trend, where the degree of positive correlation started to grow notably on December 31 (0.64), reached peak on 2020 January 20 (0.76), and then fell back (0.76 to 0.34). In other words, economic gaps matter in COVID-19 awareness, and people from lower-income locations are less aware of the pandemic than those from higher-income locations. This information gap can lead to inequalities in pandemic preparedness.

Additionally, we examined the time-evolving correlations between a series of socio-cultural and structural factors and regional awareness distributions. These factors include cultural tightness (26, 27), percentage of paddy rice (28), technology innovation index (29), proportion of illiterate population (aged 15 and above), and percentage of multiethnic households. We found significant positive and negative correlations between these factors and awareness during the cryptic transmission period, which became more pronounced as the pandemic progressed: these correlations began to strengthen notably during the beginning phase, peaked during the growth or peak phase, and weakened in the subsequent stages. Specifically, cultural tightness was positively correlated with awareness. Higher illiteracy rates and a greater proportion of multiethnic households were associated with lower awareness. In contrast, higher paddy rice percentages and a higher technological innovation index were positively correlated with awareness. See [Supplementary Material, 4.C](#) for detailed information.

These findings can be explained by cultural dimensions theory (30), tightness-looseness theory (27), and social capital theory (31). These theories highlight how inequalities in social structure, cultural diversity and resource distribution can lead to differences in information dissemination and processing, behavioral responses, and adaptability. Cultural dimensions theory suggests that cultural tightness, characterized by strict norms, can facilitate more effective information dissemination and quicker responses during pandemics (32), helping to unify public health responses



**Fig. 4.** The trends of geography-related Spearman's rank correlation coefficients: distance to epicenter (Wuhan) vs. awareness percentage (for 366 major cities); confirmed COVID-19 cases vs. awareness percentages, GDPs vs. awareness percentages, cultural tightness vs. awareness percentages, paddy rice percentages vs. awareness percentages, technology innovation indexes vs. awareness percentages, illiterate population proportions vs. awareness percentages, multiethnic household percentages vs. awareness percentages (31 provinces). Following the beginning phase, factors such as the distance from the epicenter, the proportion of the illiterate population, and the percentage of multiethnic households exhibit a negative correlation with awareness. In contrast, the number of confirmed COVID-19 cases, GDP, percentage of paddy rice, and technological innovation index show a positive correlation with awareness.



and reduce information gaps. The connection between rice farming and tight social norms (28) may explain why regions with greater paddy rice cultivation exhibit higher awareness during severe pandemic phases. Historically, the cooperative capabilities developed through rice farming foster greater collective mobilization and social cohesion in response to public health threats, consistent with tightness-looseness theory. Social capital theory emphasizes the negative impact of social inequality on pandemic responses. High illiteracy rates often indicate an unequal distribution of educational resources (33), while language barriers and cultural differences in multiethnic households can hinder information dissemination (34). Enhanced technological capabilities may improve risk perception (35) and scientific prevention behaviors (36), leading to more effective awareness. Overall, these insights suggest that in collectivist cultures with unequal social capital, pandemic awareness depends on the level of available social capital. Groups with higher social capital achieve greater awareness, while those with lower social capital face disadvantages. Cultural tightness, emphasizing strict norms and collective action, can help mitigate these disparities by promoting consistent and timely awareness. However, rising social inequalities weaken overall awareness, especially for disadvantaged groups.

## The impact of real-world events on population awareness

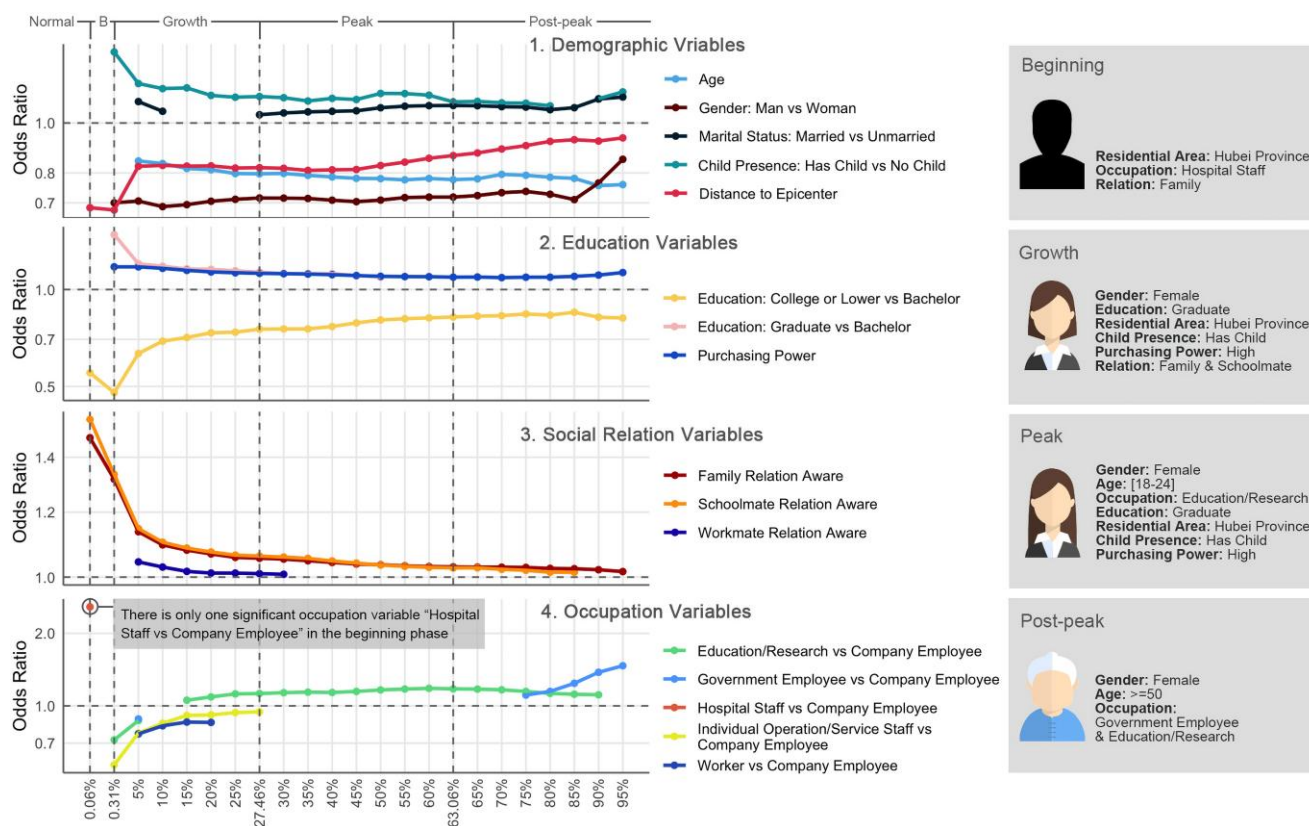
Our results hint at factors that are associated with awareness and may shape a population's reaction to an emerging pandemic. During the cryptic transmission period, several communities can advance a faster awareness response to local and public events, e.g. local news from the epicenter. When the pandemic news spread

nationally, awareness spread to all communities. Local, national, and international news can make different contributions to the diffusion of awareness. In general, the impact of national events outperforms local ones, e.g. "China NHC confirmed human-to-human transmission" (2020 January 20) is the landmark of the growth phase, while some local events alerted certain communities. Meanwhile, different awareness hysteresis effects may result from various events, e.g. after "Wuhan MHC released a briefing about the pneumonia outbreak," the aware population increased by 10% within 1.5 h, 50% after 58 h, and 100% after 9 days. Contrastingly, after "China NHC confirmed human-to-human transmission," the aware population increased by 10% just in 38 s, 50% after 4 h and 34 min, and 100% after 12 h and 26 min.

## Modeling the dynamics of public awareness

### Dynamic awareness modeling with regression

Our study focused on a descriptive analysis of the socioeconomic factors that may affect an individual's ability to make adequate preparations for their health and protection. Can pandemic awareness be meaningfully predicted from these factors? We investigate this matter by modeling individual pandemic awareness with a logistic regression model (37) where the dependent variable is an individual's Boolean awareness label set  $\{L_i^t\}$  (individual  $i$ 's awareness label, "aware" or "not aware," at time  $t$ ). We employed 11 different independent variables to predict an individual's awareness status: *Gender, Age, Occupation, Education, Distance to epicenter, Purchasing power, Child presence, Marital status, and Family/Schoolmate/Workmate awareness percentages* (see Fig. 5). More detailed feature descriptions can be found in [Supplementary](#)



**Fig. 5.** Logistic regression models (left), at different awareness percentage points, visualize the time-evolving trends of odds ratios of demographic and social relation features. X-axis is aware percentages and Y-axis is the odds ratios of variables. The typical characteristics of an aware individual (right) across the different phases.



**Material, Table S3.** Since we can condition this awareness model for each time  $t$ , it allows us to characterize an individual's evolving awareness status over time.

For the experiment, we randomly sampled 100,000 individuals (from 94,534,663 individuals) to train time-evolving logistic regression models. The remaining individuals are retained to estimate the heterogeneous social networks required for some of the independent variables (familial, educational, and professional relationships). We fit a logistic regression model for each time  $t$  when the overall aware percentage increased by 1% (from 1% to 95%) plus the time points when important real-world events occur, yielding 106 models. Each logistic regression model estimates the log-odds of an individual becoming aware of the pandemic, i.e. the outcome that  $L_i = 1$ . The results are depicted in Fig. 5. For each timestamp, only significant variables ( $P < 0.05$ ) are presented, and Y-axis represents the odds ratio (see **Supplementary Material, 5.B** for the detailed regression models and results).

The following observations can be made from the outcomes of these models. First, the geo-location, gender, child-presence, and age are the significant characteristics of an individual in the response to the emerging pandemic. Individuals close to the epicenter are able to respond effectively ( $OR_{\text{distance}} < 1$ ,  $P < 0.05$ ). The odds of being aware would significantly decrease for older individuals ( $OR_{\text{age}} < 1$ ,  $P < 0.05$ ) during the growth to post-peak phases. Second, highly educated and wealthy individuals respond to the emerging pandemic effectively. In the beginning phase (i.e. awareness percentage = 0.06%), when PPEs and other resources were still widely available, compared with the individuals with bachelor degree, the low-educated individuals (college or lower) are at a disadvantage ( $OR = 0.55$ ,  $P < 0.05$ ) to become aware of the pandemic, and this disadvantage for them continues throughout the whole observation period. Compared with the individuals with bachelor degree, individuals with PhD or master degrees show advantages in the response to the pandemic ( $OR > 1$ ,  $P < 0.05$ ) from the growth to mid-peak phases. The individuals with high purchasing power are able to respond effectively ( $OR > 1$ ,  $P < 0.05$ ) in the whole observation period. Third, compared to white-collar company employee, hospital staff can be aware quickly ( $OR = 2.58$ ,  $P < 0.05$ ) in the beginning phase; and the individuals with other occupations are at a disadvantage ( $OR < 1$ ,  $P < 0.05$ ) to become aware in the growth phase. Fourth, the family and schoolmate relations may be the critical relations for the awareness diffusion, especially in the early stages. In the beginning phase, family ( $OR_{\text{family}} = 1.48$ ,  $P < 0.001$ ) and schoolmate ( $OR_{\text{schoolmate}} = 1.56$ ,  $P < 0.001$ ) relations are both significant.

With respect to variables related to chronological "Typical Aware Individual," by leveraging significant features ( $P < 0.05$ ), we characterize the typical individual for each phase who is aware of the emerging pandemic. In the beginning phase, the aware individuals can be close to the epicenter, work in a hospital, and have aware family member(s). In the growth phase, aware individuals are more likely to be well-educated, female, has-child, close to epicenter, and high income. The aware family and schoolmate relations can be the vital factors to enhance their awareness probability. In the peak phase, well-educated, young (age  $\in [18, 24]$ ), female individuals with "education/research" occupation are more likely to be aware. In the post-peak phase, a typical aware individual can be a senior (age  $\geq 50$ ) female individual. The visualized typical aware individual can be found in Fig. 5.

## Discussion and conclusion

### Data and methodological strengths

**Early time coverage:** this study offers a comprehensive analysis of the cryptic transmission period of COVID-19, from 2019 December

1 to 2020 February 26, capturing the earliest public reactions to the pandemic. Unlike previous research that primarily focuses on the post-outbreak stage (38–40), this study provides valuable early evidence for understanding the formation of awareness and the dynamics of social behavior during the initial phase of the pandemic. **Large-scale data analysis:** utilizing data from 94 million individuals and 150 billion query and purchase records from an e-commerce platform, it employs a large-scale data analysis to reveal behavioral differences among various socioeconomic groups, offering a deeper understanding of the dynamic shifts in social inequality during a public health crisis. The data scale in this study far surpasses that of previous studies (41). **Data authenticity and reliability:** by leveraging e-commerce behavior data in its natural state, the study avoids self-report biases and observer effects. Compared with studies that rely on surveys or self-report data (39, 40, 42, 43), this approach can more accurately reflect actual decision-making behaviors of individuals. **Interdisciplinary and innovative approach:** it introduces an innovative interdisciplinary approach by integrating demographic characteristics with heterogeneous social networks and applying deep learning techniques to large-scale behavioral data, offering a distinct perspective for analyzing social dynamics and advancing data-driven methods in social science research.

### Theoretical contributions

First, this study highlights the dynamic nature of the pandemic awareness among different socioeconomic groups by analyzing mask-related search behaviors rather than actual mask-wearing. We demonstrate that the spread of pandemic awareness not only evolves over time but also can be impacted by social inequalities. This finding broadens previous research on COVID-19 inequalities (34, 44–47) and pandemic awareness (11, 48), emphasizing that public awareness is shaped by a combination of socioeconomic background, social capital distribution, and the progression of the pandemic. It enriches the theoretical understanding of how awareness and vulnerabilities shift with changing conditions, knowledge diffusion, and socioeconomic factors. Second, this study finds that populations in regions with stricter social norms and greater cultural tightness exhibit higher levels of awareness during severe phases of a pandemic's cryptic transmission period. Previous research (40, 49) has shown a positive relationship between collectivism and pandemic compliance. Our research further reveals a more nuanced relationship, showing that cultural tightness and tight social norms may not only influence compliance but also enhance awareness under conditions of uncertainty, thereby fostering more effective collective mobilization and social cohesion during public health threats. These findings extend the application of cultural dimensions theory (30) and tightness-looseness theory (27), highlighting the complex and dynamic influence of cultural factors on pandemic awareness. Third, by examining the specific roles of family, schoolmate, and workmate relations in the spread of awareness, this study provides refined insights into how relational dynamics function in the context of a pandemic. Our findings indicate that during early pandemic stages, family, and schoolmate relations were identified as significant channels for awareness diffusion. This finding challenges the traditional emphasis on broader community or institutional-level channels, suggesting that personal and direct relationships may be more effective at certain stages of pandemic response. This offers a more detailed understanding of how awareness is distributed and diffused across different social network structures.

## Practical implications

**Utilizing E-commerce data for dynamic strategy adjustment:** this study suggests that analyzing behavior data on e-commerce platforms during the early stages of a pandemic can provide real-time insights into the spread of awareness. Public health strategies should use this data to dynamically adjust their focus to increase pandemic awareness and promote preventive behaviors. **Targeted interventions for subdivision of disadvantaged communities:** public health policies should prioritize interventions tailored to disadvantaged communities, ensuring equitable access to information and resources. Deliberate attention must be given to subdivision of disadvantaged communities during each stage of a pandemic's emergence. By addressing socioeconomic risk factors and improving information and resource access for groups with lower social capital, policymakers can bridge awareness gaps and reduce disparities in pandemic outcomes, enhancing community resilience. **Optimizing communication through social networks:** public health authorities should effectively utilize social networks for information dissemination in the early stages of a pandemic and continuously adapt their strategies as the situation evolves. Early interventions can prioritize leveraging family and schoolmate networks to ensure rapid information dissemination to socially disadvantaged groups, thereby raising awareness and promoting preventive behaviors more effectively. As the pandemic evolves, these strategies can expand to broader community engagement, ensuring that measures remain adaptive and responsive to changing conditions. **Adapting public health strategies to cultural contexts:** the study's findings indicate that different cultures and social structures lead to varying responses and dynamics of awareness spread during a pandemic. Public health policies should adapt to these cultural differences across various pandemic phases to enhance effectiveness. For instance, in regions characterized by tight social norms, public health authorities may need to bolster external validation and disseminate information through trusted sources to mitigate potential delays in awareness. Additionally, leveraging cultural tightness to enhance pandemic awareness and preparedness, such as designing public health campaigns to reinforce existing norms of compliance and collective action, may improve the speed and consistency of responses.

## Conclusion

Our analysis reveals how the information about COVID-19 outbreak unfolded in China, showing a conspicuous information gap between different populations, in particular during the early stages of the pandemic. This information gap, combined with subsequent shortages of essential PPE and other supplies, could induce inequities in the health outcomes of millions. Online purchase patterns show that the most vulnerable demographics are also the last to become aware of an emerging health crisis, and therefore least favorably placed to take precautions to protect themselves, their families, and their wider social network. This study leverages a large eCommerce dataset to investigate individual search and purchase characteristics at the individual, familial, institutional, and societal social level, observing the traces of emerging awareness through a population and its (inequitable) diffusion. This study not only expands the theoretical framework concerning awareness diffusion and social inequality during pandemics but also offers practical recommendations for public health policy. It emphasizes the importance of adopting flexible and dynamic response strategies during public health crises and provides valuable insights and practical guidance on

understanding the complex influence of cultural factors and social structures on pandemic responses.

## Notes

<sup>a</sup>During this time, all 94 million individuals in our dataset were observed to be fully aware of the pandemic. The WHO–China Joint Mission on COVID-19 reported the main findings of the outbreak on 2020 February 24, which was 2 days before the end of our observation period.

<sup>b</sup><https://www.statista.com/outlook/emo/ecommerce/china>

<sup>c</sup>More detailed dataset information, privacy and ethical considerations, and machine learning inference information are available in the [Supplementary Material](#).

<sup>d</sup>Data from the normal phase can reflect background information. For instance, during the normal phase, individuals with children are more likely to search for masks compared with those without children.

## Acknowledgments

The authors thank the editor and anonymous reviewers for their valuable help and feedback, and Zekun Wang for his valuable assistance.

## Supplementary Material

[Supplementary material](#) is available at PNAS Nexus online.

## Funding

This work is supported in part by funds from the National Science Foundation (NSF, grant nos. 2122232 and 1636636), National Natural Science Foundation of China (NSFC, grant no. 72104212), and Defense Advanced Research Projects Agency (DARPA, grant no. HR001121C0168).

## Author Contributions

X.L. and Z.J. were responsible for conceptualization and methodology; Z.J., X.L., and Y. K. conducted investigation; Z.J., Y. K., and C.S. performed data curation and formal analysis; X.L., J.B., Y.Y.A., and Z.J. contributed to writing—original draft and writing—review & editing.

## Previous Presentation

These results were never previously presented.

## Preprints

There is no preprint of this article published.

## Data Availability

Due to privacy concerns and restrictions imposed by the data usage agreements, the raw data cannot be made publicly available. To support the understanding of our findings, we provide comprehensive statistical data, along with the 106 logistic regression results and related code files in publicly accessible repositories on Figshare. Daily growth rates and cumulative percentages of awareness and feature variables for 94,534,663 individuals (2019 December 1 to 2020 February 26) are available at <https://doi.org/>

10.6084/m9.figshare.26831710 and <https://doi.org/10.6084/m9.figshare.26832400>. Feature distributions for 100,000 randomly selected individuals at 106 critical time points are available as follows: for aware individuals at <https://doi.org/10.6084/m9.figshare.26863462>, for unaware individuals at <https://doi.org/10.6084/m9.figshare.26863477>. The overall feature distribution for the 100,000-individual sample is at <https://doi.org/10.6084/m9.figshare.26863618>. Details of the 106 regression models are at <https://doi.org/10.6084/m9.figshare.24131109>, and related code is at <https://doi.org/10.6084/m9.figshare.24131130>. An anonymized, de-identified, and desensitized sample dataset is available upon compliance with data management regulations and reasonable request, which should be directed to the corresponding author.

## References

- Burki T. 2020. Global shortage of personal protective equipment. *Lancet Infect Dis.* 20(7):785–786.
- Sodhi MMS, Tang CS, Willenson ET. 2023. Research opportunities in preparing supply chains of essential goods for future pandemics. *Int J Prod Res.* 61(8):2416–2431.
- Wu H-L, Huang J, Zhang CJP, Wai-Kit Ming ZH. 2020. Facemask shortage and the novel coronavirus disease (COVID-19) outbreak: reflections on public health measures. *EClinicalMedicine.* 21:100329.
- Shi W-Z, et al. 2022. Online public opinion during the first epidemic wave of COVID-19 in China based on Weibo data. *Humanit Soc Sci Commun.* 9:159.
- Valentin S, Mercier A, Lancelot R, Roche M, Arsevska E. 2021. Monitoring online media reports for early detection of unknown diseases: insight from a retrospective study of COVID-19 emergence. *Transbound Emerg Dis.* 68(3):981–986.
- Folayan MO, et al. 2023. A multi-country survey of the socio-demographic factors associated with adherence to COVID-19 preventive measures during the first wave of the COVID-19 pandemic. *BMC Public Health.* 23(1):1413.
- Laajaj R, et al. 2022. Understanding how socioeconomic inequalities drive inequalities in COVID-19 infections. *Sci Rep.* 12(1):8269.
- Suh J, Horvitz E, White RW, Althoff T. 2022. Disparate impacts on online information access during the COVID-19 pandemic. *Nat Commun.* 13(1):7094.
- Worby CJ, Chang H-H. 2020. Face mask use in the general population and optimal resource allocation during the COVID-19 pandemic. *Nat Commun.* 11(1):1–9.
- Ye Y, et al. 2022. Equitable access to COVID-19 vaccines makes a life-saving difference to all countries. *Nat Hum Behav.* 6(2):207–216.
- Funk S, Gilad E, Watkins C, Jansen VAA. 2009. The spread of awareness and its impact on epidemic outbreaks. *Proc Natl Acad Sci U S A.* 106(16):6872–6877.
- Kamalrathne T, Amaratunga D, Haigh R, Kodituwakku L. 2023. Need for effective detection and early warnings for epidemic and pandemic preparedness planning in the context of multi-hazards: lessons from the COVID-19 pandemic. *Int J Disaster Risk Reduct.* 92:103724.
- Cen Y, et al. 2019. Trust relationship prediction in Alibaba e-commerce platform. *IEEE Trans Knowl Data Eng.* 32(5):1024–1035.
- Addo PC, Jiaming F, Kulbo NB, Liangqiang L. 2020. Covid-19: fear appeal favoring purchase behavior towards personal protective equipment. *Serv Ind J.* 40(7-8):471–490.
- World Health Organization (WHO). 2020. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19).
- Huang C, et al. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 395(10223):497–506.
- Li Q, et al. 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New Eng J Med.* 382(13):1199–1207.
- Wu F, et al. 2020. A new coronavirus associated with human respiratory disease in China. *Nature.* 579(7798):265–269.
- Gozzi N, et al. 2023. Estimating the impact of COVID-19 vaccine inequities: a modeling study. *Nat Commun.* 14(1):3272.
- World Health Organization et al. *Pandemic influenza preparedness and response: a WHO guidance document.* World Health Organization, Geneva, 2009.
- Shalizi CR, Thomas AC. 2011. Homophily and contagion are generically confounded in observational social network studies. *Sociol Methods Res.* 40(2):211–239.
- Bord RJ, O'Connor RE. 1997. The gender gap in environmental attitudes: the case of perceived vulnerability to risk. *Soc Sci Q.* 78(4):830–840.
- Otterbring T, Festila A. 2022. Pandemic prevention and personality psychology: gender differences in preventive health behaviors during COVID-19 and the roles of agreeableness and conscientiousness. *J Saf Sci Resil.* 3(1):87–91.
- Stockemer D, Plank F, Niemann A. 2021. The COVID-19 pandemic and government responses: a gender perspective on differences in public opinion. *Soc Sci Q.* 102(5):2383–2393.
- Pirie W. 1988. Spearman rank correlation coefficient. *Encyclopedia Stat Sci.* 8:584–587.
- Chua RYJ, Huang KG, Jin M. 2019. Mapping cultural tightness and its links to innovation, urbanization, and happiness across 31 provinces in China. *Proc Natl Acad Sci U S A.* 116(14):6720–6725.
- Gelfand MJ, et al. 2011. Differences between tight and loose cultures: a 33-nation study. *Science.* 332(6033):1100–1104.
- Talhelm T, English AS. 2020. Historically rice-farming societies have tighter social norms in China and worldwide. *Proc Natl Acad Sci U S A.* 117(33):19816–19824.
- China Science, Technology Development Strategy Research Group, China Innovation, and University of Chinese Academy of Sciences Entrepreneurship Management Research Center. *China regional innovation capacity evaluation report: 2020.* Scientific and Technical Document Press, 2020.
- Hofstede G. 1980. Culture and organizations. *Int Stud Manag Organ.* 10(4):15–41.
- Hawe P, Shiell A. 2000. Social capital and health promotion: a review. *Soc Sci Med.* 51(6):871–885.
- Thompson FM, Brouthers KD. 2021. Digital consumer engagement: national cultural differences and cultural tightness. *J Int Mark.* 29(4):22–44.
- Street BV. 2011. Literacy inequalities in theory and practice: the power to name and define. *Int J Educ Dev.* 31(6):580–586.
- Clark E, Fredricks K, Woc-Colburn L, Bottazzi ME, Weatherhead J. 2020. Disproportionate impact of the COVID-19 pandemic on immigrant communities in the United States. *PLoS Negl Trop Dis.* 14(7):e0008484.
- Lima ML, Barnett J, Vala J. 2005. Risk perception and technological development at a societal level. *Risk Anal Int J.* 25(5):1229–1239.
- Qin H, et al. 2024. The mass public's science literacy and co-production during the COVID-19 pandemic: empirical evidence from 140 cities in China. *Humanit Soc Sci Commun.* 11(1):1–13.
- McDonald JH. *Handbook of biological statistics.* Vol. 2. Sparky House Publishing, Baltimore, MD, 2009.



- 38 Berger LM, et al. 2024. Inequality in high-cost borrowing and unemployment insurance generosity in us states during the COVID-19 pandemic. *Nat Hum Behav.* 8(9):1676–1688.
- 39 Chung RY-N, et al. 2021. Socioeconomic inequality in mental well-being associated with COVID-19 containment measures in a low-incidence asian globalized city. *Sci Rep.* 11(1):23161.
- 40 Lu JG, Jin P, English AS. 2021. Collectivism predicts mask use during COVID-19. *Proc Natl Acad Sci U S A.* 118(23):e2021793118.
- 41 Kalocsanyiova E, Essex R, Fortune V. 2023. Inequalities in COVID-19 messaging: a systematic scoping review. *Health Commun.* 38(12):2549–2558.
- 42 Damette O, Huynh TLD. 2023. Face mask is an efficient tool to fight the COVID-19 pandemic and some factors increase the probability of its adoption. *Sci Rep.* 13(1):9218.
- 43 English AS, Talhelm T, Tong R, Li X, Su Y. 2022. Historical rice farming explains faster mask use during early days of China's COVID-19 outbreak. *Curr Res Ecol Soc Psychol.* 3:100034.
- 44 Altmejd A, Östergren O, Björkegren E, Persson T. 2023. Inequality and COVID-19 in Sweden: relative risks of nine bad life events, by four social gradients, in pandemic vs. prepandemic years. *Proc Natl Acad Sci U S A.* 120(46):e2303640120.
- 45 Bathina KC, ten Thij M, Valdez D, Rutter LA, Bollen J. 2021. Declining well-being during the COVID-19 pandemic reveals us social inequities. *PLoS One.* 16(7):1–13.
- 46 Kontokosta CE, Hong B, Bonczak BJ. 2024. Socio-spatial inequality and the effects of density on COVID-19 transmission in us cities. *Nat Cities.* 1(1):83–93.
- 47 Yerkes MA, et al. 2024. Gender inequality in work location, childcare and work-life balance: phase-specific differences throughout the COVID-19 pandemic. *PLoS One.* 19(6):e0302633.
- 48 Despres C, Aguilar R, McAlister A, Ramirez AG. 2020. Communication for awareness and action on inequitable impacts of COVID-19 on latinos. *Health Promot Pract.* 21(6):859–861.
- 49 Kitayama S, Camp NP, Salvador CE. 2022. Culture and the COVID-19 pandemic: multiple mechanisms and policy implications. *Soc Issues Policy Rev.* 16(1):164–211.