

# Urbanites' mental health undermined by air pollution

Received: 20 February 2022

Accepted: 25 November 2022

Published online: 23 January 2023

Zhi Cao<sup>1,2</sup>, Jingbo Zhou<sup>2</sup>, Meng Li<sup>3</sup>, Jizhou Huang<sup>4</sup> & Dejing Dou<sup>2</sup>

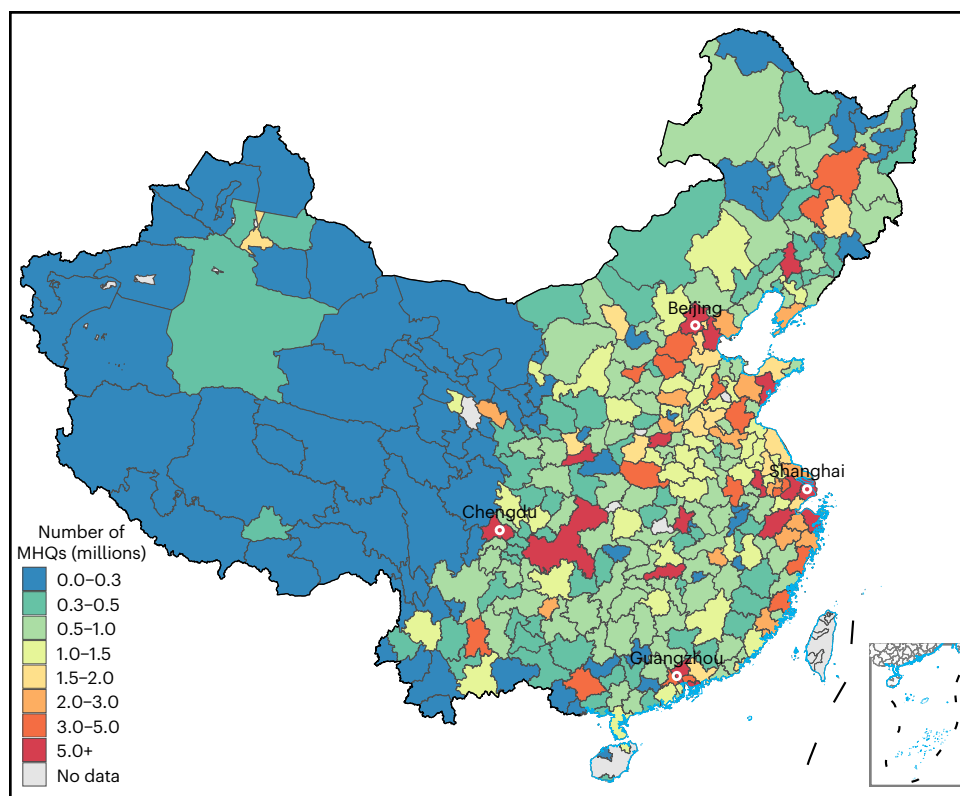
The rising mental health difficulties of the urban population in developing countries may be attributed to the high levels of air pollution. However, nationwide large-scale empirical works that examine this claim are rare. In this study, we construct a daily mental health metric using the volume of mental-health-related queries on the largest search engine in China, Baidu, to test this hypothesis. We find that air pollution causally undermines people's mental health and that this impact becomes stronger as the duration of exposure to air pollution increases. Heterogeneity analyses reveal that men, middle-aged people and married people are more vulnerable to the impact of air pollution on mental health. More importantly, the results also demonstrate that the cumulative effects of air pollution on mental health are smaller for people living in cities with a higher gross domestic product per capita, more health resources, larger areas of green land and more sports facilities. Finally, we estimate that with a one-standard-deviation increase of fine particulate matter ( $26.3 \mu\text{g m}^{-3}$ ), the number of people who suffer from mental health problems in China increases by approximately 1.15 million. Our findings provide quantitative evidence for the benefits of reducing air pollution to promote mental health and well-being.

Rapid economic development in China has been accompanied by an increase in material living standards as well as a rise in mental health difficulties, especially for urban inhabitants. According to the estimation of the World Health Organization (WHO), the number of people suffering from depression had reached 54.8 million in China (approximately 4.2% of the Chinese population) in 2015 (ref. <sup>1</sup>), and this number is increasing<sup>2</sup>. Mental disorders worsen human health<sup>3</sup>, lower productivity<sup>4</sup> and reduce quality of life<sup>5,6</sup>. They have become a major contributor to the Chinese disease burden, accounting for -3.1–7.3% of years lived with disability in 2015 (ref. <sup>1</sup>). Noticeable changes in social, economic and physical systems in China in recent decades are major factors that lead to the risks of mental disorders. Air pollution, the primary environmental challenge for China and a notable byproduct of China's industrialization<sup>7</sup>, is likely to exacerbate the risk factors for mental disorders.

Recently, a few empirical studies have shed light on the threats that air pollution poses to mental health. For example, Zhang et al.<sup>8</sup> investigated the effects of air pollution on respondents' mental health status over the previous month in China. Xue et al.<sup>9</sup> and Yang et al.<sup>10</sup> used survey data from China to estimate the impact of long-term exposure to air pollution. Newbury et al.<sup>11</sup> and Bakolis et al.<sup>12</sup> examined the association between air pollution and mental health on the basis of survey data from the United Kingdom, but the former study focused only on adolescent psychotic experiences. Some researchers have also utilized depression-related admissions to hospitals to explore the impact of air pollution on mental health<sup>13–16</sup>.

These prior studies provide an initial understanding of the relationship between air pollution and mental health. However, we still lack a large-scale nationwide quantification of the mental health risks posed by air pollution, especially with regard to the different effects with

<sup>1</sup> Business School, Sichuan University, Chengdu, China. <sup>2</sup> Business Intelligence Lab, Baidu Research, Beijing, China. <sup>3</sup> C. T. Bauer College of Business, University of Houston, Houston, TX, USA. <sup>4</sup> Baidu Inc., Beijing, China. ✉ e-mail: [zhoujingbo@baidu.com](mailto:zhoujingbo@baidu.com); [mli@bauer.uh.edu](mailto:mli@bauer.uh.edu); [huangjizhou01@baidu.com](mailto:huangjizhou01@baidu.com); [dejingdou@gmail.com](mailto:dejingdou@gmail.com)



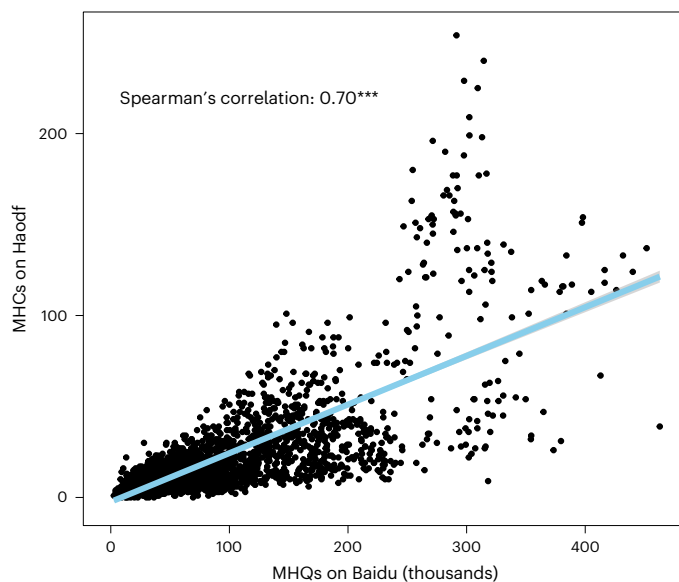
**Fig. 1 | The density distribution of MHQs in China.** The spatial distribution of geotagged search queries about mental health on Baidu. Four major cities are marked: Beijing, Shanghai, Guangzhou and Chengdu.

varied durations of pollution exposure. Previous studies employing approaches such as surveys<sup>8–10</sup> are insufficient to capture the real-time impact of air pollution on a national scale. It is unclear whether their findings hold across cities, considering the large economic disparities among different cities in China. For example, a notably negative effect between fine particulate matter ( $PM_{2.5}$ ) concentration and mental health was identified in Shijiazhuang<sup>13</sup> but not in Beijing<sup>16</sup>.

This study begins to fill this knowledge gap by utilizing a unique nationwide dataset with a fine-grained time series. Specifically, we use real-time internet search data from 1 March 2019 to 31 December 2019 across 252 cities in China to estimate the impact of short-term and long-term exposure to air pollution on urbanites' mental health at the city level. We have access to the daily queries of all users on the largest search engine in China, Baidu (Baidu.com), which allows us to obtain real-time mental-health-related queries (MHQs). Web search queries provide ample information about the interests, concerns and intentions of the overall population<sup>17</sup>, making them a valuable source of information about health trends<sup>18,19</sup>. MHQs enable us to leverage the advantages of big data to quantify the impact of air pollution on the whole population and assess the number of affected people compared with conventional approaches.

To accurately filter MHQs from daily search queries, we selected a group of search terms to capture two common mental health problems—depression and anxiety<sup>1</sup> (see the details in the Methods and Supplementary Note 1). We obtained 360 million geotagged queries related to mental health and measured people's daily mental health for each city by aggregating the MHQ data to the city level (Fig. 1). Note that the higher the volume of MHQs per capita, the worse the mental health of the population in that area (Supplementary Fig. 1).

This study answers three research questions. First, do real-time exposure and long-term exposure to air pollution have different effects on Chinese urbanites' mental health? Using the fixed effects model,



**Fig. 2 | The correlation between MHQs on Baidu and MHCs on Haodf.** Each dot represents the weekly city value of MHQs and MHCs. The blue line is the linear fitted relationship of these dots ( $R^2 = 0.55$ ), and the grey error band depicts the 95% confidence interval. \*\*\* $P < 0.001$  (two-sided test).

we quantitatively analysed the duration of exposure to air pollution that leads to a decline in people's mental health, and we show that the effects become very larger as the duration of exposure to air pollution increases (approximately six times more for 60 days of long-term exposure compared with daily exposure). Second, how does air pollution

**Table 1 | The impact of short-term exposure to air pollution on mental health**

Dependent variable: ln(number of MHQs)	Ordinary least squares		IV $d_{ij} \in (120, 300]$ km		IV $d_{ij} \in (120, 500]$ km	
	(1)	(2)	(3)	(4)	(5)	(6)
AQI	0.0016*** (0.0000)		0.0042*** (0.0000)		0.0059*** (0.0010)	
PM <sub>2.5</sub>		0.0015*** (0.0008)		0.0050*** (0.0000)		0.0069*** (0.0012)
Controls	✓	✓	✓	✓	✓	✓
City-month fixed effects	✓	✓	✓	✓	✓	✓
Date fixed effects	✓	✓	✓	✓	✓	✓
Observations	76,052	76,052	71,404	71,404	71,752	71,752
Adjusted R <sup>2</sup>	0.99	0.99	0.99	0.99	0.99	0.99

Columns 1 and 2 present the regression results from employing ordinary least squares on equation (2). Columns 3–6 present the second-stage regression results from employing two-stage least squares on equation (2), where the IV is Neighbour. Here  $d_{ij}$  is the distance between city  $i$  and grid  $j$ ;  $w_{ijt} = 1.0$  when calculating the IV (see the details in the Methods and Supplementary Note 3). Robust standard errors are clustered at the city level.  $P$  values are reported in parentheses. \*\*\* $P < 0.01$  (two-sided test).

unequally affect people's mental health for different population groups? The large-scale MHQs enable us to investigate whether the impact on mental health varies by different demographic characteristics (for example, gender, marriage and age) and socio-economic characteristics (for example, economic development, health resources and living conditions). Finally, we estimated how many people suffer from mental health issues as a result of air pollution. This estimation can shed light on the risks and losses posed by air pollution to people's mental health nationwide.

## Results

Before conducting the empirical analyses, we first adopted real mental-health-related cases (MHCs) from Haodf (Haodf.com), a leading online health care platform in China, to validate whether the MHQs are related to people's mental health in a local area (see the details in Supplementary Note 1). Figure 2 shows a highly correlated relationship between MHQs on Baidu and MHCs on Haodf (Spearman's correlation is 0.70, and  $P < 0.01$ ). To further validate the MHQs, we also employed the text of search queries—that is, we used the top ten most frequent search queries that are most likely to reflect mental health issues and examined the correlation between MHQs and these queries (see the details in Supplementary Note 1). Supplementary Fig. 3 also reveals a highly correlated relationship (Spearman's correlation is 0.90, and  $P < 0.01$ ). We thus found that MHQs can capture the mental health status of the population in a city.

### Short-term exposure to air pollution

We first estimated the short-term effect of air pollution on urbanites' mental health using equation (2). As shown in columns 1 and 2 of Table 1, there is a significantly positive association between air pollution (air quality index (AQI) and PM<sub>2.5</sub>) and MHQs, suggesting that people's mental health status declines when they are temporarily exposed to air pollution. The coefficient in column 2 indicates that a one-standard-deviation increase in PM<sub>2.5</sub> concentration (that is, 26.3  $\mu\text{g m}^{-3}$ ) is associated with a 0.0033-standard-deviation increase in search queries about mental health on Baidu. The magnitude is practically notable given the large volume of search queries. For example, a 0.0033-standard-deviation increase translates into an increase of approximately 1.44 million MHQs during our study period. While this query volume is not equal to the actual increase in the number of people with mental health problems, it should be highly correlated<sup>18</sup>. According to our statistics, each user on Baidu searched for mental health information 7.8 times on average, and the average ratio of the

located queries was 46.41% during our sample period. Therefore, an increase in MHQs of 1.44 million means an increase of approximately 0.398 million people (that is, 1.44 million/7.8/0.4641) who suffer from mental health problems due to air pollution. It is worth noting that the true value should be higher than this number because not all people look up mental-health-related information through the search engine. The results for other pollutants (that is, PM<sub>10</sub>, SO<sub>2</sub>, CO, NO<sub>2</sub> and O<sub>3</sub>) are reported in Supplementary Table 5.

In addition to air pollution, extreme weather might also affect people's mental health<sup>20</sup>. Accordingly, we compared the impact of air pollution with that of an extreme storm, Typhoon Lekima, which had the most severe influence in China in 2019, to help us understand the effect size. We found that the impact of air pollution on mental health issues was smaller than that of Typhoon Lekima (for example, the impact of PM<sub>2.5</sub> was that a one standard deviation increase in PM<sub>2.5</sub> concentration will increase the volume of MHQs by 0.0067 standard deviation and the one of Typhoon Lekima was a 0.0141 standard deviation increase in the search volume of MHQs; Supplementary Note 2 and Supplementary Table 6). Nonetheless, severe air pollution occurs frequently in China, unlike extreme storms. For example, 53.0% of cities' daily PM<sub>2.5</sub> concentrations exceeded the WHO limitation in the sample period.

The relationship between air pollution and mental health might be spurious as a result of omitted variables that vary with days on the city level. For instance, traffic congestion may affect a city's air pollution and exacerbate people's emotions. The results would be biased if such omitted variables were not controlled in the model specification. We leveraged an instrumental variable (IV) to address the endogeneity issues caused by omitted variables, which helped us identify the causal effect of air pollution on mental health.

We constructed the IV—Neighbour—by combining the wind direction with the air pollution level of neighbouring areas following prior research<sup>21</sup> (Methods). The insight behind this method is that the formation and dissipation of air pollution are heavily affected by meteorological conditions<sup>22</sup>; the wind can bring exogenous variation to a city's air pollution level by blowing pollutant emissions from neighbouring regions to the city<sup>21,23</sup>. The neighbour is an ideal IV as it is unlikely to influence the social and economic activities of a city, except for varying the city's air pollution. This estimation also addresses the classic measurement error issue in using station-based data<sup>21</sup>. The IV results reveal that the coefficients of AQI and PM<sub>2.5</sub> remain positive and statistically significant (columns 3–6 in Table 1; see the full results in Supplementary Table 7). The results are robust to altering the weight and the range in calculating the IV (see the details in Supplementary

Note 3 and Supplementary Table 8). This demonstrates the causal effect of air pollution on people's mental health.

In addition to the IV estimation, our baseline results are robust to various checks (Supplementary Note 3 and Supplementary Table 11). First, Supplementary Table 11a documents similar results with the alternative measurement of the dependent variable, where we accounted for the different population sizes of regions by using the ratio of MHQs instead of the natural logarithm of MHQs. Second, we included day-of-week fixed effects and quarter fixed effects to absorb the possible interference from different weekdays and seasons (Supplementary Table 11b). Third, we removed duplicate queries of each user on each day; that is, we counted only one effective query if a user searched many MHQs in one day (Supplementary Table 11c). Fourth, we used a two-way cluster standard error—that is, simultaneously clustering by cities and days—to capture the unspecified correlations between observations for different cities on the same day (Supplementary Table 11d). Finally, we controlled for the number of Baidu users to mitigate the concern that the usage ratios of Baidu in different regions might affect the MHQs (Supplementary Table 11e). We also conducted a placebo test using false search queries (see the details in Supplementary Note 3 and Supplementary Table 13).

### Long-term exposure to air pollution

We employed equation (3) to estimate the long-term effects of air pollution. We controlled for transitory exposure to air pollution in this model to evaluate the relative importance of transitory and cumulative effects. Table 2 reports the results using four windows of air pollution exposure: 7-day, 14-day, 30-day and 60-day exposures. As shown in Table 2, the impact of air pollution becomes stronger as the duration of exposure increases, even when controlling for the real-time effects of air pollution. The coefficient in column 4 suggests that a one-standard-deviation increase in  $PM_{2.5}$  concentration during the past 60 days is related to a 0.0206-standard-deviation increase in MHQs (over six times the short-term effect of  $PM_{2.5}$ , 0.0033). These findings reveal that the proportion of people who suffer from mental health problems strikingly increases if the air quality in the city remains at a poor level for a long time. The results are quite similar when using the AQI to reflect air pollution (Supplementary Table 23).

### Heterogeneity analyses

We further examined how air pollution impacts mental health among different demographic and socio-economic groups. We considered four demographic characteristics: gender, education, age and marriage. First, we compared the effects of air pollution on the male and female subsamples (Supplementary Table 14). Figure 3a shows the estimated results coupled with 95% confidence intervals. As shown in Fig. 3a, the effect on mental health becomes larger for both men and women as the duration of exposure to air pollution increases. The results also reveal that although the impact of short-term exposure to air pollution is not significantly different between men and women, men are more vulnerable than women as the duration of exposure increases. This finding is similar to the study by Zhang et al.<sup>24</sup>, who found that the effect of air pollution on cognitive performance is larger for men than for women.

We repeated our analysis to explore heterogeneous effects on people with different educational attainments. Figure 3b and Supplementary Table 15 display the estimated coefficients for two subsamples with different educational attainments—that is, high school or below versus college or above. As shown in Fig. 3b, the effects of air pollution are significant for both groups, but there are no significant differences in the impacts on people with high or low levels of educational attainment.

To learn how air pollution affects mental health for different age cohorts, we estimated equation (3) on the basis of five cohorts: 18–24, 25–34, 35–44, 45–54 and 55–64 years. Supplementary Table 16 reports the numerical results, and Fig. 3c shows the graphical results. In general, we found that the effects (both short-term and long-term effects) of

**Table 2 | The impact of long-term exposure to air pollution on mental health**

Dependent variable: $\ln(\text{number of MHQs})$	7-day (1)	14-day (2)	30-day (3)	60-day (4)
$\frac{1}{k} \sum_{n=0}^{k-1} PM_{2.5}$	0.0060*** (0.0000)	0.0086*** (0.0000)	0.0091*** (0.0000)	0.0137*** (0.0000)
$PM_{2.5}$	−0.0001 (0.7779)	0.0005 (0.1527)	0.0011** (0.0187)	0.0012** (0.0110)
Controls	✓	✓	✓	✓
City-month fixed effects	✓	✓	✓	✓
Date fixed effects	✓	✓	✓	✓
Observations	76,052	76,052	76,052	76,052

This table presents the regression results from employing ordinary least squares on equation (3). Robust standard errors are clustered at the city level. *P* values are reported in parentheses. \*\**P* < 0.05; \*\*\**P* < 0.01 (two-sided test).

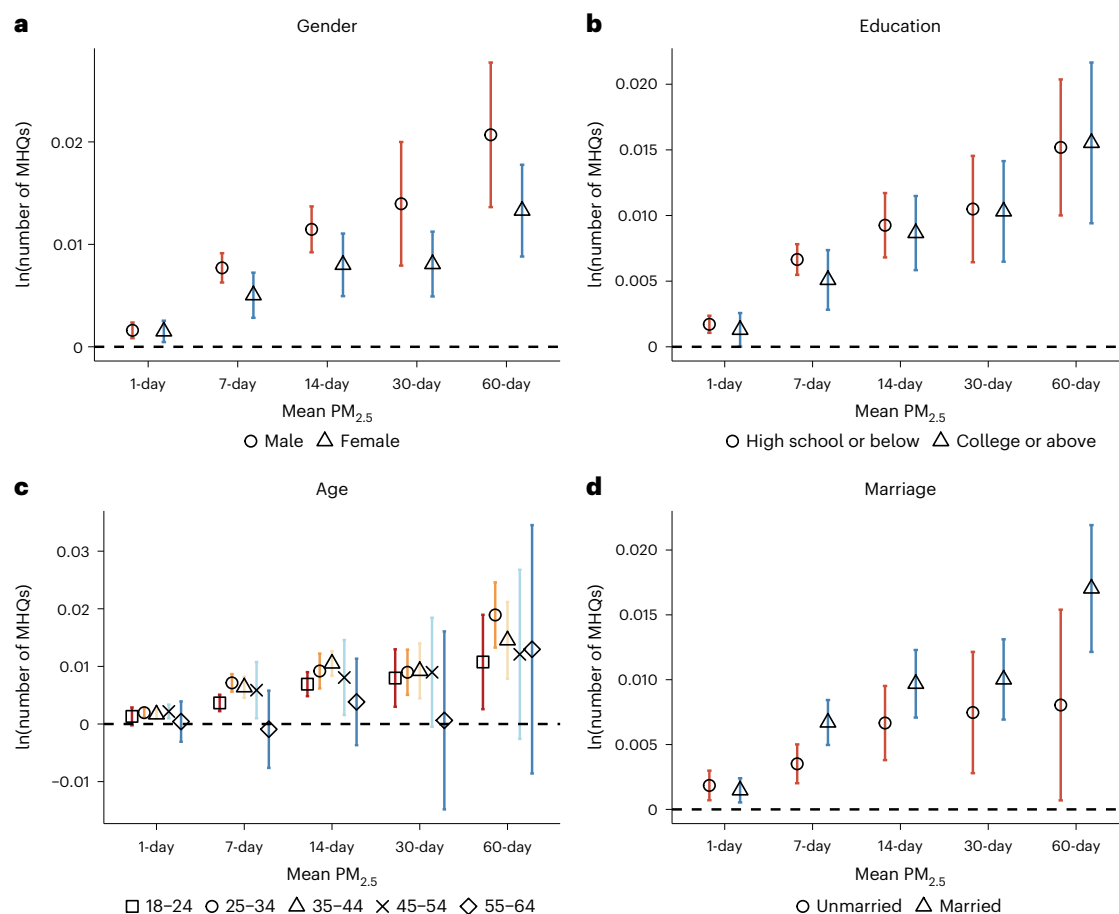
air pollution on mental health present an inverted-U shape as people age. Specifically, compared with the youngest age cohort (that is, 18–24) and the oldest age cohort (that is, 55–64), the effect on mental health caused by air pollution is more pronounced for the middle-aged cohorts (that is, 25–54).

We also compared the married group with the unmarried group to examine whether the effects of exposure to air pollution are different. Figure 3d presents the graphical results (the numerical results are in Supplementary Table 17), indicating that the impact of air pollution on mental health is stronger for married people than for unmarried people, especially when they are exposed to air pollution over the long term.

We next explored how the effects of exposure to air pollution vary with the socio-economic characteristics of each city. We examined the impact of heterogeneity with respect to economic development, health resources, living conditions and sports facilities using gross domestic product (GDP) per capita, the number of hospitals, the area of green land and the number of gyms. Figure 4 reports the graphical results (Supplementary Tables 18–21 report the numerical regression results). We found that the effects of long-term exposure to air pollution are smaller for people living in cities with higher GDP per capita, more health resources, larger areas of green land and more sports facilities.

Prior literature has demonstrated that impoverished countries (that is, low- and middle-income countries) suffer from more mental disorders<sup>25,26</sup>. Our study reveals that this inequality remains within a large country such as China. People living in cities with lower economic development suffer from more mental health problems caused by air pollution. An environment with a higher quality of living could improve people's subjective well-being, thus alleviating the effect of air pollution. More health resources could enable people to conveniently receive professional treatment for mental disorders and reduce mental health problems. More accessible green spaces in urban areas are associated with decreased mental illnesses (for example, anxiety disorder)<sup>27</sup>, thereby mitigating the effect of air pollution. Physical exercise is also beneficial to promoting people's mental health<sup>28,29</sup>. Hence, the impact of air pollution could be weakened in cities with more sports facilities, as citizens have more opportunities to access them. Our findings imply that improving the living conditions and public social welfare of cities can effectively reduce the negative effects of air pollution on urbanites' mental health. We also obtain consistent results using the AQI to measure the air quality of each city (see the details in Supplementary Note 4, Supplementary Figs. 6 and 7, and Supplementary Tables 24–31).





**Fig. 3 | The heterogeneous effects of air pollution on mental health by gender, education, age and marriage. a**, The heterogeneous effects of  $PM_{2.5}$  on mental health for the male subsample ( $N = 75,758$ ) and female subsample ( $N = 75,787$ ). **b**, The heterogeneous effects of  $PM_{2.5}$  on mental health for the high school or below subsample ( $N = 75,772$ ) and the college or above subsample ( $N = 75,775$ ). **c**, The heterogeneous effects of  $PM_{2.5}$  on mental health for different age cohorts:

18–24 ( $N = 75,727$ ), 25–34 ( $N = 75,757$ ), 35–44 ( $N = 75,578$ ), 45–54 ( $N = 74,676$ ) and 55–64 ( $N = 59,142$ ). **d**, The heterogeneous effects of  $PM_{2.5}$  on mental health for the married group ( $N = 75,793$ ) and the unmarried group ( $N = 75,701$ ). The points represent the estimated coefficients of the cumulative effects of  $PM_{2.5}$  (that is,  $\frac{1}{k} \sum_{n=0}^{k-1} PM_{2.5}$ ) from Supplementary Tables 14–17; the error bars depict the 95% confidence intervals.

### Additional analysis

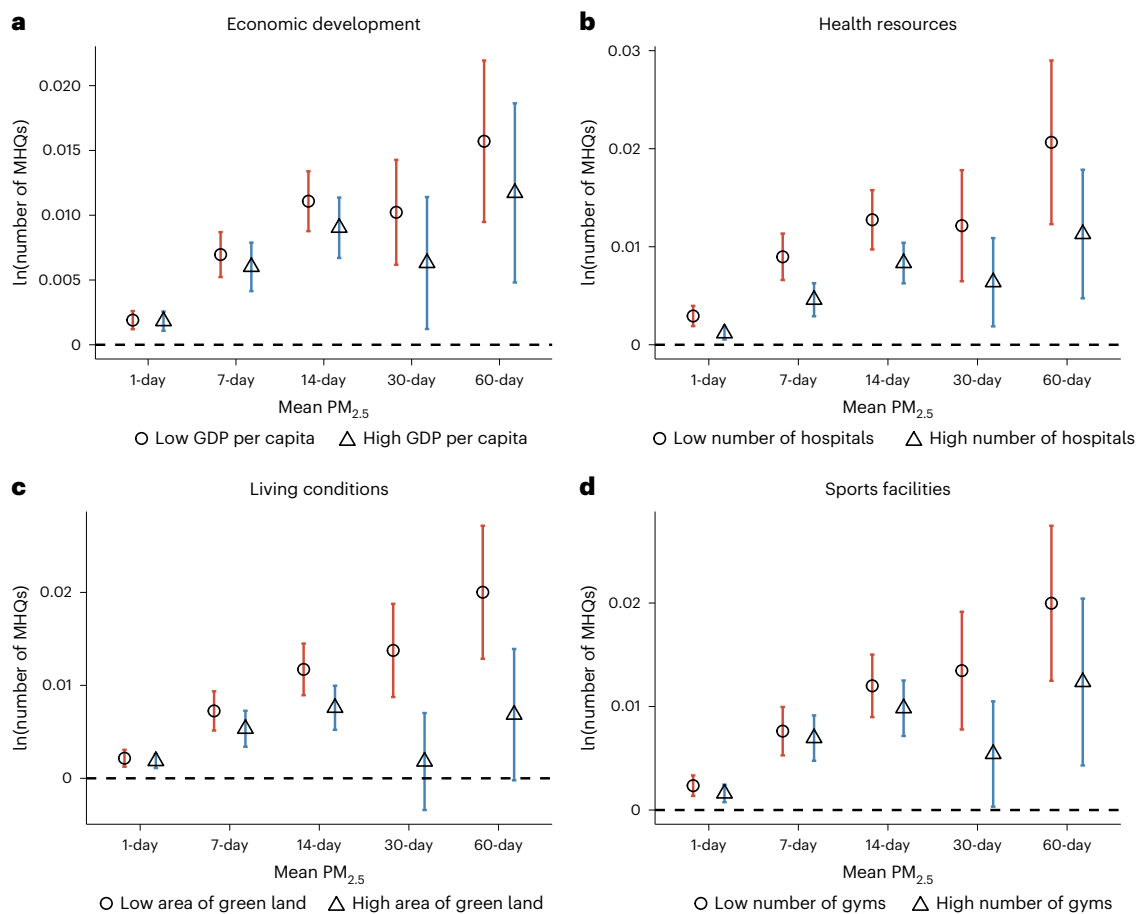
To further investigate whether air pollution has different effects on depression and anxiety (the two most common mental illnesses<sup>1</sup>), we separated the MHQs into two subsamples: a depression-related query subsample and an anxiety-related query subsample (Supplementary Table 2). We then estimated the impacts on depression and anxiety on the basis of equation (3). Supplementary Table 22 presents the results. It shows that air pollution has statistically significant impacts on both depression- and anxiety-related queries. However, the likelihood of suffering depression is higher under long-term exposure to air pollution (see the details in Supplementary Note 5).

### Discussion

Air pollution is a severe environmental problem around the world, especially in developing countries such as China. Previous studies have documented the adverse effects of air pollution on people's health<sup>30–34</sup>, but research on mental health using nationwide large-scale data remains scarce. This study investigates the short-term and long-term impacts of air pollution on urbanites' mental health by leveraging national real-time internet search data in China. Our findings suggest that both short-term and long-term exposures to air pollution exacerbate mental health problems, manifested by the statistically significant increase in search queries about these problems. The effects become larger as the duration of exposure to air pollution increases.

Furthermore, the heterogeneity analyses reveal that the impact of air pollution is stronger for the male group, the middle-aged group and married people, especially when the duration of exposure to air pollution increases. We did not find statistically significant differences for people with different educational attainment. More importantly, the results also demonstrate that the cumulative effects of air pollution on mental health are smaller for people living in cities with higher GDP per capita, more health resources, larger areas of green land and more sports facilities.

A nationwide study is needed to evaluate the representative exposure–response links between air pollution and mental health given the research limitations of prior studies. Our findings confirm the causal effects of air pollution on human mental health. The extensive heterogeneity analyses in this study also provide a comprehensive understanding of how air pollution affects people's mental health under various demographic and socio-economic conditions. Moreover, this study helps estimate how many people suffer from mental health problems as a result of air pollution. Given the market share of the Baidu search engine (69.5%)<sup>35</sup> and the percentage of internet users who used the search engine service in China in 2019 (approximately 49.6%)<sup>36</sup>, this study found that a one-standard-deviation increase in  $PM_{2.5}$  (that is,  $26.3 \mu g m^{-3}$ ) results in approximately 1.15 million people with mental health problems (that is, 0.398 million/0.695/0.496). Overall, our findings provide evidence that campaigns to mitigate air



**Fig. 4 | The heterogeneous effects of air pollution on mental health by economic development, health resources, living conditions and sports facilities.** **a**, The heterogeneous effects of economic development measured by GDP per capita ( $N = 35,068$  in the high group;  $N = 34,675$  in the low group). **b**, The heterogeneous effects of health resources measured by the number of hospitals ( $N = 32,976$  in the high group;  $N = 32,541$  in the low group). **c**, The heterogeneous effects of living conditions measured by the areas of green land ( $N = 32,356$  in the

high group;  $N = 32,243$  in the low group). **d**, The heterogeneous effects of cultural and sports facilities measured by the number of gyms ( $N = 29,330$  in the high group;  $N = 28,919$  in the low group). The points represent the estimated coefficients of the cumulative effects of  $PM_{2.5}$  (that is,  $\frac{1}{k} \sum_{n=0}^{k-1} PM_{2.5}$ ) from Supplementary Tables 18–21; the error bars depict the 95% confidence intervals. We used the median values to separate the high group from the low group for each pair of heterogeneity analyses.

pollution, despite potentially large and unavoidable control costs, can bring extra benefits by improving people's mental health.

## Methods

### Baidu search data

According to the report of the China Internet Network Information Centre<sup>36</sup>, the total number of internet users using the Baidu search engine reached 694 million in 2019, where Baidu accounts for approximately 70% of the market share<sup>35</sup>. The search queries on Baidu are thus representative and able to reflect the activities of most populations.

We partnered with Baidu to obtain all search queries after 1 March 2019 on the Baidu search engine. To filter the MHQs from daily search queries, we selected a group of search terms related to two mental health problems—depression and anxiety. According to WHO<sup>1</sup>, depression and anxiety are the two most common mental disorders and are highly prevalent in the overall population. We obtained seven keywords for depression and eight keywords for anxiety referring to the definitions of WHO<sup>1</sup> (analyses based on each keyword are presented in Supplementary Note 6). We then translated these keywords to Chinese following the processes in Supplementary Fig. 2 to build the corresponding Chinese search keywords (Supplementary Note 1 and Supplementary Table 2). On the basis of these keywords, we constructed

a regular expression to filter the MHQs from daily search queries on Baidu:

$$\text{Regex} = [\mathbf{A}] * ![\mathbf{B}], \quad (1)$$

where **A** is a set of selected keywords representing depression and anxiety and **B** is a set of words that are used to exclude the irrelevant queries. The elements included in **A** and **B** are reported in Supplementary Table 2. The rule of this regular expression is that a query is an effective MHQ if at least one keyword in set **A** is included in the query and all keywords in set **B** do not occur in this query.

Using the filtering rule, we obtained 360 million geotagged search queries (with a 95.1% positive ratio) related to mental health from 252 cities from 1 March 2019 to 31 December 2019. We adopted this period to avoid confounding factors from COVID-19 because many cities were locked down at the start of 2020, which might affect people's mental health<sup>37</sup>.

### Online doctor consultation data

In recent years, online doctor consultation has been an important way of broadening the channels available to patients. The online consultation records the actual mental health cases. Accordingly, we collected online doctor consultation data from a leading online health care

platform in China, Haodf, to validate the association between MHQs on Baidu and MHCs on Haodf. The MHCs were selected on the basis of the disease classification on Haodf and included typical mental disorders, such as depressive disorder, anxiety disorder and insomnia (Supplementary Note 1). We then aggregated the data to the city level.

### Air quality data

We obtained air quality data from China's Ministry of Ecology and Environment, which provides real-time monitoring data covering all prefectural cities. The original dataset included hourly readings of the AQI and specific pollutants, including PM<sub>2.5</sub> concentrations, from 1,605 monitoring stations. We aggregated the station-hour-level air pollution data to city-day-level data by averaging all station data within the corresponding city each day.

### Weather data

We considered five common weather variables: temperature, sunshine, humidity, precipitation and wind speed. The weather data were collected from the China Meteorological Data Service Centre, which provides daily meteorological data from 699 meteorological stations covering most cities. We calculated the distance between each city and all meteorological stations. We then used the meteorological data from the nearest station as a proxy for the city weather conditions. We excluded cities for which the distance to the nearest meteorological station is more than 40 km. This reduced the number of our sample cities to 252. We also varied this threshold to verify the robustness of our results (Supplementary Note 3 and Supplementary Table 12).

### Demographic and socio-economic data

To explore heterogeneity, we assembled demographic and socio-economic data. The demographic data were at the individual level and were provided by the Baidu User Profile Platform, which integrates the user information from over 55 Baidu products (apps) and creates a comprehensive profile for more than 90% of individual users in Baidu. The profile data were generated by collecting the users' actual information and predicting unknown information on the basis of comprehensive machine learning algorithms. We collected each individual's demographic information from the platform and constructed mental health measures for different groups (such as male versus female) for each city. The socio-economic data were at the city level and came from the *2019 China City Statistical Yearbook* compiled by the Chinese National Bureau of Statistics. The yearbook provides a series of city information, including GDP per capita, number of hospitals and area of green land. We collected each city's socio-economic information from the yearbook and divided the sample cities into high- and low-value groups on the basis of the median value of each characteristic.

### Baseline regression model

We used a fixed effect panel model to estimate the effects of air pollution on mental health. First, we estimated the short-term effects of air pollution on the basis of the following econometric specification:

$$\ln(\text{MHQs})_{it} = \alpha_0 + \alpha_1 \text{AirPollution}_{it} + \alpha_2 \mathbf{X}_{it} + \mu_{i,m} + \lambda_t + \epsilon_{it}, \quad (2)$$

where  $\ln(\text{MHQs})_{it}$  is the natural logarithm of the number of MHQs in city  $i$  on date  $t$ ;  $\text{AirPollution}_{it}$  represents the pollution level in city  $i$  on date  $t$ , measured by AQI and PM<sub>2.5</sub> concentration;  $\mathbf{X}_{it}$  is a vector of weather controls;  $\mu_{i,m}$  indicates city-month fixed effects, which can not only control for time-invariant confounders but also absorb time-varying characteristics for each city (such as local business cycle dynamics); and  $\lambda_t$  indicates date fixed effects, accounting for common shocks for all cities in a given day. The coefficient  $\alpha_1$  reflects the transitory impact of air pollution on mental health and is expected to be positive.

Next, we estimated the long-term effects of air pollution on mental health by focusing on the primary pollutant—PM<sub>2.5</sub>:

$$\ln(\text{MHQs})_{it} = \alpha_0 + \alpha_1 \text{PM}_{2.5it} + \alpha_2 \cdot \frac{1}{k} \sum_{n=0}^{k-1} \text{PM}_{2.5i,t-n} + \alpha_3 \mathbf{X}_{it} + \mu_{i,m} + \lambda_t + \epsilon_{it}. \quad (3)$$

Here PM<sub>2.5it</sub> represents the contemporaneous level of air pollution in city  $i$  on date  $t$ , and  $\frac{1}{k} \sum_{n=0}^{k-1} \text{PM}_{2.5i,t-n}$  is the mean level of air pollution in the past  $k$  days, which measures cumulative exposure. The control variables are the same as in equation (2). The coefficient  $\alpha_1$  reflects the short-term impact of air pollution on mental health, and  $\alpha_2$  reflects the long-term exposure on mental health. Controlling for these two variables in a model can help us evaluate the relative importance of transitory and accumulative effects in the association between air pollution and mental health.

### IV estimation

We introduced an IV and employed the two-stage-least-square approach to address the potential endogeneity issues in equation (2). In this study, we followed the idea of Zheng et al.<sup>21</sup> by utilizing “cross-boundary air pollution flows” to construct the IV.

To build the cross-boundary spillover measure of local air pollution, we procured a dataset from Tracking Air Pollution in China (<http://tapdata.org>), which provides real-time PM<sub>2.5</sub> data with a spatial resolution of 10 km × 10 km (in a regular grid of 0.1° × 0.1°)<sup>38</sup>. We also used the PM<sub>2.5</sub> data from the monitoring stations to construct the IV and obtained very similar results (see the details in Supplementary Note 3 and Supplementary Tables 9 and 10). Using the data, we constructed the IV:

$$\text{Neighbour}_{it} = \frac{1}{k} \sum_{j=1}^k w_{ijt} \times \text{GridPollution}_{jt}, d_{ij} \in (120, 300] \text{ km}, \quad (4)$$

where  $\text{GridPollution}_{jt}$  represents grid cell  $j$ 's pollution level on date  $t$  measured by PM<sub>2.5</sub> concentration;  $w_{ijt}$  represents grid cell  $j$ 's weight, which depends on the wind direction of city  $i$  on date  $t$  and the relative direction of grid cell  $j$  to city  $i$ ; and  $d_{ij}$  is the distance between city  $i$  and grid cell  $j$ . The weights of upwind grid cells are larger than those of grid cells in other directions, and the weights of downwind grid cells are set to zero because their impacts on city  $i$ 's air pollution level are minimal.  $\text{Neighbour}_{it}$  therefore measures how city  $i$ 's air pollution level on date  $t$  is affected by the PM<sub>2.5</sub> concentration from nearby grid cells on the same day.

To minimize the interference from agglomeration economies on a group of cities, which may produce correlations between the IV and city  $i$ 's economic activities<sup>21</sup>, we excluded all the neighbouring grid cells for which the distance to city  $i$  is within 120 km. The correlation between grid cell  $j$ 's PM<sub>2.5</sub> concentration and city  $i$ 's air pollution level should be small if the distance between city  $i$  and grid cell  $j$  is long. We thus also excluded grid cells outside 300 km from city  $i$  given that 90% of the sample days' wind speeds are smaller than 3.4 m s<sup>-1</sup> (that is, 312 km d<sup>-1</sup>—the greatest distance the wind can travel per day).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The air quality data are available from <https://air.cnemc.cn>. The weather data are available from <http://data.cma.cn>. The socio-economic data are available from <https://data.cnki.net/area/Yearbook/Single/N2020050229>. The spatial PM<sub>2.5</sub> data that were used to generate the IV are available from <http://tapdata.org>. The online doctor consultation data are available from <https://open.haodf.com/opendata/home>. After anonymization and aggregation, there are no privacy issues for the aggregated mental health query data and related demographic data

for academic purposes. However, these data are still sensitive for the respective communities since they were extracted from user-generated data. Therefore, the aggregated mental health query data and related demographic data that support the findings of this study are available from the corresponding author Jingbo Zhou upon reasonable request.

## Code availability

All R code for this paper is available at Zenodo: <https://doi.org/10.5281/zenodo.7352528> (ref. <sup>39</sup>).

## References

1. *Depression and Other Common Mental Disorders: Global Health Estimates* (World Health Organization, 2017).
2. Huang, Y. et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. *Lancet Psychiatry* **6**, 211–224 (2019).
3. Moussavi, S. et al. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* **370**, 851–858 (2007).
4. Stewart, W. F., Ricci, J. A., Chee, E., Hahn, S. R. & Morganstein, D. Cost of lost productive work time among US workers with depression. *JAMA* **289**, 3135–3144 (2003).
5. Whiteford, H. A. et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575–1586 (2013).
6. Deryugina, T., Heutel, G., Miller, N. H., Molitor, D. & Reif, J. The mortality and medical costs of air pollution: evidence from changes in wind direction. *Am. Econ. Rev.* **109**, 4178–4219 (2019).
7. *Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease* (World Health Organization, 2016).
8. Zhang, X., Zhang, X. & Chen, X. Happiness in the air: how does a dirty sky affect mental health and subjective well-being? *J. Environ. Econ. Manage.* **85**, 81–94 (2017).
9. Xue, T., Zhu, T., Zheng, Y. & Zhang, Q. Declines in mental health associated with air pollution and temperature variability in China. *Nat. Commun.* **10**, 2165 (2019).
10. Yang, Z. et al. Air pollution and mental health: the moderator effect of health behaviors. *Environ. Res. Lett.* **16**, 044005 (2021).
11. Newbury, J. B. et al. Association of air pollution exposure with psychotic experiences during adolescence. *JAMA Psychiatry* **76**, 614–623 (2019).
12. Bakolis, I. et al. Mental health consequences of urban air pollution: prospective population-based longitudinal survey. *Soc. Psychiatry Psychiatr. Epidemiol.* **56**, 1587–1599 (2021).
13. Song, J. et al. Acute effects of ambient particulate matter pollution on hospital admissions for mental and behavioral disorders: a time-series study in Shijiazhuang, China. *Sci. Total. Environ.* **636**, 205–211 (2018).
14. Lowe, S. R., Wang, C., Ma, Y. & Chen, K. Particulate matter pollution and risk of outpatient visits for psychological diseases in Nanjing, China. *Environ. Res.* **193**, 110601 (2021).
15. Chen, C. et al. Ambient air pollution and daily hospital admissions for mental disorders in Shanghai, China. *Sci. Total. Environ.* **613**, 324–330 (2018).
16. Gao, Q., Xu, Q., Guo, X., Fan, H. & Zhu, H. Particulate matter air pollution associated with hospital admissions for mental disorders: a time-series study in Beijing, China. *Eur. Psychiatry* **44**, 68–75 (2017).
17. Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M. & Watts, D. J. Predicting consumer behavior with web search. *Proc. Natl Acad. Sci. USA* **107**, 17486–17490 (2010).
18. Ginsberg, J. et al. Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012–1014 (2009).
19. Liu, K. et al. Using Baidu Search Index to predict dengue outbreak in China. *Sci. Rep.* **6**, 38040 (2016).
20. Cunsolo, A. & Ellis, N. R. Ecological grief as a mental health response to climate change-related loss. *Nat. Clim. Change* **8**, 275–281 (2018).
21. Zheng, S., Wang, J., Sun, C., Zhang, X. & Kahn, M. E. Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nat. Hum. Behav.* **3**, 237–243 (2019).
22. Arain, M. et al. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmos. Environ.* **41**, 3453–3464 (2007).
23. Zhang, Q. et al. Transboundary health impacts of transported global air pollution and international trade. *Nature* **543**, 705–709 (2017).
24. Zhang, X., Chen, X. & Zhang, X. The impact of exposure to air pollution on cognitive performance. *Proc. Natl Acad. Sci. USA* **115**, 9193–9197 (2018).
25. Funk, M., Drew, N. & Knapp, M. Mental health, poverty and development. *J. Public Ment. Health* **11**, 116–185 (2012).
26. Knapp, M. et al. Economic barriers to better mental health practice and policy. *Health Policy Plan.* **21**, 157–170 (2006).
27. Nutsford, D., Pearson, A. & Kingham, S. An ecological study investigating the association between access to urban green space and mental health. *Public Health* **127**, 1005–1011 (2013).
28. *Promoting Mental Health: Concepts, Emerging Evidence, Practice—Summary Report* (World Health Organization, 2004).
29. Ten Have, M., de Graaf, R. & Monshouwer, K. Physical exercise in adults and mental health status: findings from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *J. Psychosom. Res.* **71**, 342–348 (2011).
30. Chen, Y., Ebenstein, A., Greenstone, M. & Li, H. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proc. Natl Acad. Sci. USA* **110**, 12936–12941 (2013).
31. Ebenstein, A., Fan, M., Greenstone, M., He, G. & Zhou, M. New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. *Proc. Natl Acad. Sci. USA* **114**, 10384–10389 (2017).
32. Brunekreef, B. & Holgate, S. T. Air pollution and health. *Lancet* **360**, 1233–1242 (2002).
33. Neidell, M. Information, avoidance behavior, and health the effect of ozone on asthma hospitalizations. *J. Hum. Resour.* **44**, 450–478 (2009).
34. Zhang, L. et al. Air pollution-induced missed abortion risk for pregnancies. *Nat. Sustain.* **2**, 1011–1017 (2019).
35. *Search Engine Market Share China* (StatCounter, accessed 6 June 2022); <https://gs.statcounter.com/search-engine-market-share/all/china/2019>
36. *The 44th Statistical Report on Internet Development in China* (China Internet Network Information Centre, accessed 6 June 2022); <http://www.cac.gov.cn/pdf/20190829/44.pdf>
37. Brodeur, A., Clark, A. E., Fleche, S. & Powdthavee, N. COVID-19, lockdowns and well-being: evidence from Google Trends. *J. Public Econ.* **193**, 104346 (2021).
38. Geng, G. et al. Tracking air pollution in China: near real-time PM<sub>2.5</sub> retrievals from multisource data fusion. *Environ. Sci. Technol.* **55**, 12106–12115 (2021).
39. Zhi, C., Jingbo, Z., Meng, L., Jizhou, H. & Dejing, D. Urbanites' mental health undermined by air pollution (code). Zenodo <https://doi.org/10.5281/zenodo.7352528> (2022).

## Acknowledgements

We thank all the members of the Business Intelligence Lab for their support and encouragement. We thank Yixiong Xiao and Aijuan Han for helpful discussions and suggestions. This work is partially supported by the National Engineering Research Center of Deep Learning Technology and Application.



## Author contributions

This work was accomplished when Z.C. was working as a student intern in Baidu Research supervised by J.B.Z. J.B.Z. came up with the idea of using internet search data to estimate the impact of air pollution on negative emotions. Z.C., J.B.Z., M.L., J.Z.H. and D.J.D. jointly conceived the study. Z.C., J.B.Z. and M.L. designed the analysed model and experiments. Z.C. and J.B.Z. collected and processed the search data. Z.C. implemented the analysed model and ran the model. Z.C., J.B.Z. and M.L. performed the analysis. J.Z.H. advised on the data processing and literature review. D.J.D. advised on the experimental design. Z.C., J.B.Z. and M.L. wrote the paper. J.B.Z., J.Z.H. and D.J.D. managed the project. All authors contributed to review and editing.

## Competing interests

J.B.Z., J.Z.H. and D.J.D. are currently affiliated with Baidu. Z.C. was affiliated with Baidu Research as a student intern when conducting this research. M.L. declares no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41893-022-01032-1>.

**Correspondence and requests for materials** should be addressed to Jingbo Zhou, Meng Li, Jizhou Huang or Dejing Dou.

**Peer review information** *Nature Sustainability* thanks Hong Wu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated  |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection All collected data are processed using custom written Python code to generate samples used to regression analyses (code available upon request).

Data analysis R version 4.0.3 is used to conduct regressions and plot figures (code available upon request).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The air quality data are available from <https://air.cnemc.cn>. The weather data are available from <http://data.cma.cn>. The socioeconomic data are available from <https://data.cnki.net/area/Yearbook/Single/N2020050229>. The spatial PM2.5 data that are used to generate the instrumental variable are available from <http://tapdata.org>. The online doctor consultation data are available from <https://open.haodf.com/opendata/home>. Other data that support the findings of this study, including the mental health-related queries data and related demographic data, are available from the corresponding author upon reasonable request.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study has constructed a daily city-level mental health metric based on the volume of mental health-related queries (MHQs) on the largest search engine in China---Baidu, to test how air pollution affects mental health of urbanites. We use the real-time Internet search data from March 1, 2019 to December 31, 2019 across 252 cities in China to estimate this relationship. We further investigate whether the impact is heterogeneous considering the different demographic and socio-economic factors.
Research sample	The sample consists of all people's search queries on Baidu.com across 252 cities in China. As Baidu.com is the biggest search engine in China, accounting for about 70% of the market share, the search behavior on Baidu.com can accurately reflect the mental health status of people in a city.
Sampling strategy	The sample size, 252 cities, is selected based on the data availability. We excluded cities where the distance to the nearest meteorological station is more than 40 km.
Data collection	Mental health-related queries data are collected and filtered from Baidu search engine. Air pollution data are collected from China's Ministry of Ecology and Environment. Weather data are collected from the China Meteorological Data Service Center. The demographic and socio-economic data are collected from the 2019 China City Statistical Yearbook.
Timing and spatial scale	The sample period is from March 1, 2019 to December 31, 2019. We select March 1, 2019 as the start date of the sample because this is the earliest date that we are permitted to access the search query data; The selection of December 31, 2019 is to avoid the confounding factors from the COVID-19, because one-third of Chinese cities were locked down to prevent the escalation of COVID-19 virus transmission at the start of 2020, which might affect people's mental health.
Data exclusions	There are no data excluded in the analysis.
Reproducibility	1. Using instrument variable estimation to check the robustness of the results. 2. Using alternative measurement for the dependent variable and rerun regression. 3. Removing duplicated queries and re-analyzing.
Randomization	We control for the city-month fixed effects, date fixed effects, and weather-related variables, including temperature, sunshine, humidity, precipitation, and wind speed.
Blinding	No blinding is performed in this study. This is because the analyses in this study do not involve any procedures that need blinding.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging