The Association Between PM$_{2.5}$ and Depression in China

Guangye He$^1$, Yunsong Chen$^2$, Senhu Wang$^3$, Yiqun Dong$^1$, Guodong Ju$^1$, and Buwei Chen$^4$

Abstract

While China has been experiencing unprecedented economic growth, depression is becoming one of the most striking social and mental health problems in recent years. Such a paradox to progress may partially be due to the notoriously poor air quality of the country. To verify this argument, we constructed an index of the prevalence of depression (IPD) using internet search query volumes in Baidu to proxy the potential depression and examined how IPD is associated with PM$_{2.5}$, the major air pollutant in China. Our results from 2-way fixed effects models reveal that a 100 $\mu$g/m$^3$ increase in previous week’s PM$_{2.5}$ in a city is significantly associated with a 0.279 increase in its IPD, comparable to 7.34 hours decrease in weekly daylight, and such relationship is particularly pronounced in the spring and summer and in East and South areas. Our findings of large-scale pattern suggest that PM$_{2.5}$ at current levels in China poses serious mental health risks.

Keywords

PM$_{2.5}$, depression, China, internet search query volumes, Baidu index

Air pollution, as an environmental hazard, imposes a variety of adverse effects on individuals’ health. Exposure to air pollution is positively associated with the incidence of stroke,$^1$ respiratory diseases,$^{2-4}$ and lung and cardiovascular dysfunction.$^5-10$ Long-term exposure to air pollution even leads to higher rates of mortality.$^{11-14}$

As Ambient Air Pollution, a 2016 global health report published by the World Health Organization (WHO)$^{15}$ reported, in 2012 alone, ambient air pollution was complicit in 3 million deaths and 85 million disability-adjusted life years.$^{16}$ Foremost among the pollutants associated with these mortality and morbidity numbers were airborne particulate matter known as PM$_{2.5}$. As a mixture of fine particles and liquid droplets, PM$_{2.5}$ is a particular danger because it is easily sucked into the lungs, where it can create a devastating impact on an individual’s health.

Aside from the role PM$_{2.5}$ plays in physical health, previous studies have touched on the potential relationship between PM$_{2.5}$ and individuals’ mental health. Depression, one of the leading cause of the disability-adjusted life years, is closely associated with physical illness and decreased quality of life.$^{17-19}$ and may also be associated with PM$_{2.5}$. However, most of the existing studies focus on high-income countries, such as Canada, Europe, and United States; few have examined the situation in low-income countries, such as China (except 2 papers.$^{24,26}$)

In the past 2 decades, China has experienced unprecedented economic growth. However, the health and longevity of the Chinese people have not improved as a result.$^{27}$ It is argued that economic growth has had the opposite effect as the suicide rate and the incidence of mental illness have rapidly increased.$^{28}$ This paradox of progress may be due, in part, to the country’s notoriously poor air quality,$^{29}$ where PM$_{2.5}$ could...
be the culprit behind hazy skies. According to the latest PM$_{2.5}$ concentration data, on April 17, 2016, the annual average PM$_{2.5}$ concentration in 210 cities in China was in the range of 11 to 128 $\mu$g m$^{-3}$, around 99% of cities surpassed 15 $\mu$g m$^{-3}$, the maximum limits of annual average PM$_{2.5}$ concentration set by WHO.$^{30}$

In this study, we attempt to examine the potential PM$_{2.5}$–depression link. We constructed a weekly index of the prevalence of depression (IPD) in 282 major cities in China using internet searches for depression-related terms on Baidu, the most widely used search engine in China. Note that there is unwarranted stigmatization associated with depression. Such stigma makes individuals embarrassed to confess their true feeling, and even avoid reaching out for help, leading to a high level of underreporting.$^{31,32}$ Internet search provides a way out, as internet searching is largely anonymous and confidential, without the presence of third party, and individuals are more likely to reflect the true feeling. Similar methods have been used in various studies to construct measures that cannot be captured in more traditional ways.$^{33-38}$

**Methods**

**Data**

The data prepared for this analysis are from multiple sources. We constructed city-level IPD, and baseline web search index using search index extracted from Baidu, the most widely used internet search engine in China. We used Python 2.7.11 to gather search data from the web and chose “week” as the time unit.

City-level weekly PM$_{2.5}$ was collected from China’s Air Quality Online Monitoring and Analysis Platform (https://www.aqistudy.cn/historydata). To rule out the potential confounding effect of weather, we collected the relevant meteorological information from the Chinese website Historical Weather (http://www.tianqihoubao.com/lishi) and the websites of Sunrise and Sunset in the website of Convenience Inquiry (https://richurimo.51240.com.zhongguo_richurimo).

**Index of Prevalence of Depression**

The dependent variable is the city-level IPD. To identify queries for depression, we searched for the 4 most common names for depressive disorders: depression (yi yu zheng), melancholia (you yu zheng), postpartum depression (chan hou yi yu zheng), and obsessive-compulsive disorder (OCD, qiang po yi yu zheng), as up to 75% of individuals with OCD have major depression.$^{39}$ We divided results into 52 weeks in 2017 and conducted a principal component analysis (PCA) to extract the largest component from the 4 items to serve as IPD (see Table 1). Similar results are shown when summing up the Z-score of the 4 time series as a proxy of IPD.

## Table 1. Factor Loadings of Depression Search Index in Baidu.$^a$

| Variables                                      | Factor | Uniqueness |
|------------------------------------------------|--------|------------|
| Depression                                     | 0.92   | 0.16       |
| Melancholia                                    | 0.83   | 0.30       |
| Obsessive-compulsive disorder (OCD)            | 0.88   | 0.22       |
| Postpartum depression                          | 0.82   | 0.32       |
| Eigenva1ue                                     | 3.00   |            |
| Variance, %                                    | 74.98  |            |
| Cronbach x                                     | 0.79   |            |

$^a$Principal component factor analysis. KMO = 0.83. Bartlett’s test of sphericity = 2.16e-05 (6), P < .001.

## Table 2. Descriptive Statistics of Major Variables.

| Variable | Description | Mean  | SD   |
|----------|-------------|-------|------|
| IPD      | Index of the prevalence of depression measured by search frequency of major depression-related words in Baidu using PCA method (log value) | 13.795 | 11.746 |
| PM$_{2.5}$ | Weekly averaged PM$_{2.5}$ 100 mcg/m$^3$ | 0.824 | 0.349 |
| Internet | Internet usage measured by the frequency of surfing 11 major websites | 0.018 | 1.009 |
| Daylen   | Weekly averaged day length multiplied by the weekly ratio of sunny day | 3.298 | 3.038 |
| Tem      | Weekly averaged day mean temperature | 16.113 | 10.509 |
| Rain     | Weekly ratio of rainy days | 0.267 | 0.242 |
| Wind     | Weekly averaged wind speed | 13.795 | 11.746 |

**Control Variables**

We control for several variables to take care of potential confounding effects and present the relevant descriptive statistics in Table 2. First, city-level internet usage. We average city-level weekly data from Baidu for 11 most popular portals (ie, Baidu, We-chat, QQ, Taobao, Alipay, Sina, Sohu, Netease, Tencent, Chinese Central Television, and Bullet strain) and conducted a PCA to extract 1 factor from these items.

Second, city-level weather conditions. They are (1) weekly daylight (Daylen) (weekly averaged day length multiplied by the weekly ratio of sunny days), (2) weekly averaged day mean temperature (Tem), (3) weekly ratio of rainy days (Rain), and (4) weekly averaged wind speed (Wind). All the information were collected from the Chinese website Historical Weather (http://www.tianqihoubao.com/lishi) and the websites of Sunrise and Sunset in the website of Convenience Inquiry (https://richurimo.51240.com.zhongguo_richurimo).

To deal with the temporal effects, we also controlled for week number (centered) and its squared term to capture the potential nonlinearity.

**Model**

To examine how PM$_{2.5}$ might affect depression in China, we employed 2-way fixed effects model that is easy to rule out confounding effects rising from time-invariant, city-level
factors. We use the lagged values ($t - 1$) of all explanatory variables to account for the potential time delay. The formula is written as follows:

$$IPD_i = \beta_1 PM_{2.5_i, t-1} + \beta_2 Internet_{i, t-1} + \beta_3 Rain_{i, t-1} + \beta_4 Wind_{i, t-1} + \beta_5 Daylen_{i, t-1} + \beta_6 Week_i + \beta_7 Sqweek_i + \epsilon_i + \mu_{it}$$

In Equation 1, $IPD_i$ is the dependent variable, denoting the level of prevalence of depression in city $i$ week $t$, while $IPD_{i, (t-1)}$ is its 1-week lagged value. Moreover, $\epsilon_i$ is the city-level, time-constant error term, and $\mu_{it}$ is other city-level, time-varying error term.

**Results**

**Distribution of Weekly PM$_{2.5}$ and IPD**

Figure 1 presents the distribution of weekly average PM$_{2.5}$ in 282 cities in China. As shown, central China and some parts of northeast and northwest China with higher proportion of heavy industry are the most polluted areas as evidenced by the highest levels of PM$_{2.5}$. In contrast, western China, Inner Mongolia, and some parts of northeast and south China have relatively low levels of air pollution.

The distribution of IPD is visualized in Figure 2. As shown, individuals living in northwest China, especially Tibet, are the least likely to search for depression-related terms, followed by individuals from northern China, and then from the south and some parts of central China. Such distribution seems in line with the regional differences in socioeconomic development, internet accessibility, and air pollution.

**Results From 2-Way Fixed Effects Model**

Table 3 reports the results of a 2-way fixed-effects model. As Model 1 demonstrates, PM$_{2.5}$ is positively and significantly associated with IPD. Specifically, a 100 $\mu g/m^3$ increase in the previous week’s PM$_{2.5}$ in a city is significantly associated with a 0.279 increase in its IPD ($P < .01$), suggesting a positive relationship between PM$_{2.5}$ and depression. Consistent with previous research, weekly daylight length ($P < .001$) and wind speed ($P < .001$) are significantly and negatively correlated with depression level. Specifically, a 1-unit increase in daylight length and wind speed level is associated with 0.038 and 0.378 increases of IPD, respectively.

As the contributions of particular components to particulate matter vary by season, we further conduct by-season analysis, presented in models 2 to 5. Results showed that PM$_{2.5}$ is positively associated with depression searches in all seasons but is more pronounced in spring ($P < .001$) and summer ($P < .01$), partly of reason is higher temperatures are often associated with a faster dissipation of airborne particles.

China covers a massive amount of land area, and PM$_{2.5}$ patterns vary according to the region for reasons of geographical features, concentrations of industry, and regional idiosyncrasies; we further fitted models by region. As shown in Table 4, PM$_{2.5}$ is associated with depression searches in all Chinese regions except the west, where both PM$_{2.5}$ and depression
Figure 2. Depression-related online search volumes in 282 Chinese cities (2017).

Table 3. Two-Way, Fixed Effects Models Predicting Depression by Seasons.a

|                | (1) | (2) | (3) | (4) | (5) |
|----------------|-----|-----|-----|-----|-----|
|                | All | Spring | Summer | Autumn | Winter |
| Lagged PM₂.⁵  | 0.279b | 0.520c | 0.670b | 0.310 | 0.036 |
| (0.089)       | (0.129) | (0.239) | (0.184) | (0.190) |
| Lagged Internet| 3.822c | 2.520b | 1.237b | 0.719 | 4.142c |
| (0.626)       | (0.846) | (0.391) | (0.763) | (1.224) |
| Lagged Tem    | −0.007 | −0.022 | −0.056b | −0.047c | 0.159c |
| (0.006)       | (0.014) | (0.020) | (0.016) | (0.016) |
| Lagged Rain   | 0.172 | 0.071 | 0.096 | −0.052 | 1.564c |
| (0.109)       | (0.194) | (0.178) | (0.176) | (0.381) |
| Lagged Daylen | −0.378c | −0.094 | −0.414 | 0.030 | −0.809 |
| (0.108)       | (0.144) | (0.214) | (0.233) | (0.497) |
| Lagged Wind   | −0.038c | 0.025 | −0.033d | −0.020 | −0.067b |
| (0.008)       | (0.014) | (0.013) | (0.018) | (0.024) |
| Week          | 0.042b | −0.239b | 0.195c | 0.113 | 0.038c |
| (0.003)       | (0.005) | (0.002) | (0.007) | (0.004) |
| Week squared  | −0.001c | −0.011c | −0.017c | −0.009b | 0.007c |
| (0.000)       | (0.002) | (0.004) | (0.003) | (0.000) |
| Constant      | 15.14c | 12.640c | 16.354c | 15.419c | 12.021c |
| (0.398)       | (0.610) | (0.876) | (0.913) | (1.726) |
| Number of cities | 282 | 277 | 277 | 282 | 282 |
| Observations  | 14 066 | 3589 | 3587 | 3607 | 3283 |
| R²            | .157 | .022 | .083 | .071 | .256 |

aStandard errors in parentheses.
bP < .01.
cP < .001.
dP < .05.

Table 4. Two-Way, Fixed Effects Models Predicting Depression by Regions.a

|                | (1) | (2) | (3) | (4) | (5) |
|----------------|-----|-----|-----|-----|-----|
|                | North | East | Central | South | West |
| Lagged PM₂.⁵  | 0.196 | 0.782c | 0.278 | 1.470c | −0.220 |
| (0.117)       | (0.291) | (0.260) | (0.322) | (0.154) |
| Lagged Internet| 1.914 | 3.462b | 6.476c | 4.890c | 3.918c |
| (0.993)       | (1.282) | (2.753) | (1.302) | (0.237) |
| Lagged Tem    | 0.031c | −0.016 | −0.002 | 0.045d | 0.003 |
| (0.011)       | (0.024) | (0.017) | (0.020) | (0.013) |
| Lagged Rain   | −0.190 | 1.041b | 0.448 | 0.412c | 0.056 |
| (0.252)       | (0.304) | (0.403) | (0.199) | (0.193) |
| Lagged Wind   | −0.467c | −0.357 | −0.987d | −0.241 | −0.573 |
| (0.184)       | (0.240) | (0.459) | (0.234) | (0.338) |
| Lagged Daylen | −0.026 | −0.019 | −0.038 | −0.124c | −0.031d |
| (0.016)       | (0.022) | (0.023) | (0.030) | (0.015) |
| Week          | 0.023c | 0.054c | 0.042c | 0.044c | 0.035c |
| (0.006)       | (0.010) | (0.008) | (0.007) | (0.005) |
| Week squared  | 0.002c | −0.002c | 0.000 | 0.000 |
| (0.001)       | (0.001) | (0.001) | (0.001) | (0.001) |
| Constant      | 13.262c | 18.450c | 16.750c | 14.326c | 12.599c |
| (0.714)       | (0.928) | (1.501) | (0.912) | (1.271) |
| City number   | 64 | 58 | 52 | 46 | 62 |
| Observations  | 3246 | 2912 | 2512 | 2282 | 3114 |
| R²            | .085 | .199 | .161 | .210 | .211 |

aStandard errors in parentheses.
bP < .01.
cP < .001.
dP < .05.
searches are low. Specifically, the positive association between PM$_{2.5}$ and depression searches is statistically significant for the east ($P < .01$) and the south ($P < .001$), where socioeconomic development level is higher than the rest of the region. Such a pattern can also be taken as a manifestation of the paradox of progress in China. That is, regions enjoy the benefits of economic gain and, meanwhile, suffer from increased depression resulting from air pollution.

Discussion

While China’s economy has boomed in the past 2 decades, the quality of life of Chinese people has not. On the contrary, there is an increased risk of depression. Such a paradox to progress might be partly attributable to air pollution. In this article, we examined whether PM$_{2.5}$, one of the most dangerous and heatedly discussed sources of air pollution, is associated with depression. We constructed a city-level weekly IPD using online query volume data from Baidu rather than self-reported depression to proximate the latent depression.

Our findings provide suggestive evidence on the source of China’s paradox of progress that economic success has not translated into a better quality of life for all. We recommend that future studies identify the sources of depression and the associated mechanisms that lead to this paradox and that they inform policy with the aim of improving the quality of life for all Chinese citizens.

Limitations

This study had some limitations. First, the valid finding of this research is premised upon the fact that no short-term social, economic, or political changes influence the PM$_{2.5}$ level and searches for depression-related terms. We did not verify these short-term changes; thus, we cannot rule out the possibility that there have been some that affected air quality. Even if there were, changes in air quality often take relatively long periods of time to be observed. In this vein, the biased estimation is less of a concern in this analysis. Second, different from most of the extant studies that involve rigorous analyses at the individual level, our estimation of depression index is conducted based on less explored macro-level and the identification of PM2.5 as one potential risk factor is negatively associated with implied depression; however, what drives such relationships remains unsolved, making the city-level pattern less intuitive. Note that due to difficulties of accessing data at macro-level and the wide concern of ecological fallacy, macro-level analyses remain rare in public health field, making the development of macro-theory stagnated. This analysis, although unsound, is a critical interrogation of available big data to harness their strength at the macro-level and from a macro perspective. It, to the least degree, can be taken as the first step to tap into the exploration of macro-theory.

Conclusion

To examine the possible correlation between PM$_{2.5}$ and depression, we matched IPD with city-level weekly PM$_{2.5}$ and weather condition variables. The results show that 1-week lag of PM$_{2.5}$ can positively predict the IPD of the following week, holding constant weekly internet usage and various meteorological factors, and such effects are more prominent in spring and summer and in the East and South.

Author Contributions

G.H., Y.C., S.W., and Y.D. developed the ideas, analyzed the data, and drafted the paper. G.J and B.C collected the data and revised the paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by the Major Project of The National Social Science Fund of China (No. 19ZDA149).

ORCID iD

Guangye He https://orcid.org/0000-0002-1777-0779

References

1. Wellenius GA, Schwartz J, Mittleman MA. Air pollution and hospital admissions for ischemic and hemorrhagic stroke among Medicare beneficiaries. *Stroke* 2005;36(12):2549-2553.
2. Adam M, Schikowski T, Carsin AE, et al. Adult lung function and long-term air pollution exposure. ESCAPE: a multi-centre cohort study and meta-analysis. *Eur Respir J*. 2015;45(1):38-50. doi:10.1183/09031936.00130014
3. Kim HJ, Choi MG, Park MK, Seo YR. Predictive and Prognostic Biomarkers of Respiratory Diseases due to Particulate Matter Exposure. *J Cancer Prev*. 2017;22(2):6-15.
4. Schikowski T, Adam M, Marcon A, et al. Association of ambient air pollution with the prevalence and incidence of COPD. *Eur Respir J*. 2014;44(3):614-626.
5. Gallagher P, Lazarus W, Shapouri H, Conway R, Bachewe F, Fischer A. Cardiovascular disease—Risk benefits of clean fuel technology and policy: a statistical analysis. *Energy Policy*. 2010;38(2):1210-1222.
6. Beelen R, Stafoggia M, Raaschou-Nielsen O, et al. Long-term exposure to air pollution and cardiovascular mortality: an analysis of 22 European cohorts. *Epidemiology*. 2014;25(3):368-378.
7. Peters A, Pope CA. Cardiopulmonary mortality and air pollution. *Lancet*. 2002;360(9341):1184-1185.
8. Pope CA, Burnett RT, Thurston GD, et al. Cardiovascular mortality and long-term exposure to particulate air pollution—Epidemiological evidence of general pathophysiological pathways of disease. *Circulation*. 2004;109(1):71-77.
9. Sorensen M, Hoffmann B, Hvidberg M, et al. Long-term exposure to traffic-related air pollution associated with blood pressure and mortality due to cardiovascular disease. *Epidemiology*. 2002;13(4):407-413.
self-reported hypertension in a Danish cohort. Environ. Health Perspect. 2012;120(3):418-424.
10. Wang Y, Eliot MN, Koutrakis P, et al. Ambient air pollution and depressive symptoms in older adults: results from the MOBILIZE Boston Study. Environ Health Perspect. 2014;122(6):553-558.
11. 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 U S A. 2013;110(32):12936-12941.
12. Ebenstein A, Fan M, Greenstone M, He G, Yin P, Zhou M. High stakes energy and environmental problems in developing countries. Am Econ Rev. 2017;105:226-231.
13. He G, Fan M, Zhou M. The effect of air pollution on mortality in China: evidence from the 2008 Beijing Olympic Games. J Env Econ Management. 2016;79:18-39.
14. Heinrich J, Thiering E, Rzehak P, et al. Long-term exposure to NO2 and PM10 and all-cause and cause-specific mortality in a prospective cohort of women. Occup Environ Med. 2013;70(3):70179-70186.
15. World Health Organization. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease. WHO Document Production Services; 2016.
16. Murray CJL. Quantifying the burden of disease: the technical basis for disability-adjusted life years. Bull World Health Organ. 1994;72(3):429-445.
17. Rumsfeld JS, Ho PM. Depression and cardiovascular disease: a call for recognition. Circulation. 2005;111(3):250-253.
18. Hare DL, Toukhasti SR, Johansson P, Jaarsma T. Depression and cardiovascular disease: a clinical review. Eur Heart J. 2014;35(21):1365-1372.
19. Ruo B, Rumsfeld JS, Hlatky MA, Liu H, Browner WS, Whooley MA. Depressive symptoms and health-related quality of life: the heart and soul study. JAMA. 2003;290(1):215-221.
20. Buoli M, Grassi S, Caldirola A, et al. Is there a link between air pollution and mental disorder? Environ Int. 2018;118:154-168.
21. Kim K, Lim Y, Bae HJ, Kim M, Jung K, Hong Y. Long-term fine particulate matter exposure and major depressive disorder in a community-based urban cohort. Environ Health Perspect. 2016;124(10):1547-1553.
22. Vert C, Sánchez-Benavides G, Martínez D, et al. Effect of long-term exposure to air pollution on anxiety and depression in adults: a cross-sectional study. Int J Hyg Env Heal. 2017;220(6):1074-1080.
23. Wang R, Xue D, Liu Y, Liu P, Chen H. The relationship between air pollution and depression in China: is neighbourhood social capital protective? Int J Environ Res Public Health. 2018;15(6):1160.