Caring about the polluted attention: evidence from the inconsistent response existing in web-based survey

Mingying Zhu, Yifan Kang & Juan Liu

Published online: 18 Jul 2022.

Submit your article to this journal ↗

Article views: 511

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Caring about the polluted attention: evidence from the inconsistent response existing in web-based survey

Mingying Zhu[a,b] (iD), Yifan Kang[c] and Juan Liu[d]

[a]Deparment of Industrial Economics, School of Economics, Nanjing University, Nanjing, China; [b]Hopkins-Nanjing Center for Chinese and American Studies, Nanjing, China; [c]School of Finance and Taxation, Shandong University of Finance and Economics, Jinan, China; [d]Department of Human Resource Management, Wuxi University, Wuxi, China

ABSTRACT

Paying attention is one of the most important brain-based cognitive skills. The detrimental effect of air pollution on brain function has been largely studied by way of experiments in Neuroscience. In this paper, we provide a new way to study the impact of air pollution - in particular PM2.5 - on brain attention through observing the inattentive responding of the participants in online surveys. We construct the inconsistent index and connect it to the individual's exposure to air pollution. The preferred estimate reveals that a one standard deviation of PM2.5 (around 30 $\mu g/m^3$) increases the likelihood of inconsistency by 2.25%. Further heterogeneous analysis shows that air pollution appears to have larger effects on those young groups with high education but low income, especially for students, and the number is 5.36%. This study explores the impact of air pollution on brain health from a new perspective, and also provides a mechanism to support the recent study about air pollution and productivity.

1. Introduction

This article attempts to explore the possible link between air pollution and people's brain attention by observing the inconsistent responses when filling out the web-based survey. After controlling for all the other confounding factors, the inconsistent index that changes with air pollution could be inferred to be driven by the polluted attention of the participants. Paying attention is the first step of learning and sustained concentration is essential in the workplace. Exploring the underlying environmental factors that affect people's attention not only helps to improve brain health in its own right, but also contributes to the existing research of air pollution and work performance.

Brain's attention, related to perception and memory, is one of the most important brain-based cognitive skills (Zhang, 2019). There is a growing body of cross-sectional

CONTACT Mingying Zhu ✉ zhumy@nju.edu.cn

literature that recognises the association between air pollutants and brain function. The common pollutants in the atmosphere mainly include dust or inhalable particulate matter (PM), sulfur dioxide (SO2), nitrogen oxides (NO2), carbon monoxide (CO) and Ozone (O3), among which PM is the most harmful component to brain health (Ao et al., 2021; Peeples, 2020; Younan et al., 2020). PM2.5 with particle size less than 2.5 microns, acting as a vector, is more likely to be attached by toxic and harmful substances (such as heavy metals, microorganisms, etc.).

Neurotoxicologist Deborah Cory-Slechta indicated that tiny particles could be carried to the brain through the olfactory nerve (Elder et al., 2006).[1] A recent study by Peeples (2020) suggested that particulate pollutants could affect the brain's function through the bloodstream, even without directly entering the brain. Cory-Slechta's team studied the newborn mice exposed to air pollution in short term, and found a large amount of inflammation existing in all of their brain regions. Moreover, they also reported that the white matter in the corpus callosum was significantly reduced, and the mice appeared to show autism, attention deficit disorder and schizophrenia (Jew et al., 2019). Kilian and Kitazawa (2018) reviewed the literature from epidemiological and animal studies, and provided a large amount of evidence supporting the neurotoxic effects of air pollutants. In this study, we provide a new way to study the impact of air pollution on brain's attention by observing the inattentive responding of the participants in online surveys.

Online survey has been one of the most popular methods to realize data collection and information processing attributable to its convenience, low cost, speed and efficiency (Cobanoglu et al., 2001; Rice et al., 2017). Since network is widely distributed, web-based surveys break the restrictions of time and space, and effectively increase the survey response rates compared to the traditional methods, like mail-in surveys and phone interviews. However, when collecting data from anonymous Internet users, the reliability and consistency of the responses have been the biggest challenges (Tuten et al., 2002).

Due to the lack of guidance and assistance from the professional investigators, online participants are more likely to be inattentive and make careless mistakes in uncontrolled settings (Cheung et al., 2017). According to the identification method from Meade and Craig (2012), there were nearly 10% to 20% of the participants exhibiting the careless behaviour when filling out the online surveys, which was also referred to as insufficient effort responding (Bowling et al., 2016). Brühlmann et al. (2020) conducted a Latent Profile Analysis and revealed that almost 45.9% of the respondents showed careless behaviour to a certain extent. This kind of error, caused by the participants' insufficient attention, could directly reduce the quality of online survey data, even irrelevant to the content of the questionnaire.

The inattentive or careless responding of participants in online surveys has been extensively studied in the fields of psychology and sociology (Curran, 2016; Maniaci & Rogge, 2014). Most of the existing research pays attention to its causes, impact on data accuracy, and how to identify and prevent the problem (Brühlmann et al., 2020; Kam & Meyer, 2015). In this paper, we focus on the external environmental factors, and provide the first evidence that the elevated air pollution increases the inconsistent response existing in web-based surveys. The underlying mechanism behind the results

could be inferred that air pollution threatens brain health through physiological pathways and reduce the attention of the participants, so that people fail to make sufficient efforts to the responding work (Jew et al., 2019).

The inconsistent response is the most representative mistake due to inattentiveness, that is people offering the different responses when answering two very similar questions. There is a large volume of research developing the psychometric tests to identify the inattentive responses in the surveys, like the Personality Assessment Inventory (PAI) developed by Leslie Morey at the Texas A&M University (Morey & Lowmaster, 2010) and the Minnesota Multiphasic Personality Inventory (MMPI) developed by Starke R. Hathaway and J. C. McKinley at the University of Minnesota (Butcher, 2010). Maniaci and Rogge (2014) mainly introduced two approaches of the infrequency scales and the inconsistency scales, and also stated their effectiveness at measuring the people's inattentive responses.[2]

In this paper, we follow the approach introduced by Maniaci and Rogge (2014) and take advantage of the unique feature existing in the Netizen Ideology Survey launched in 2014, 2015 and 2017, where the pair of identical questions or contents are presented in different ways. Based on the attitude differences in responses of the participants, we construct the inconsistent index which is a binary variable, with 0 representing the consistency and 1 for the existence of the inconsistent behaviours. Correspondingly, the individual's exposure to air pollution is matched to every respondent according to the disclosed IP address and their responding time. We implement the probit model to regress the binary inconsistent index on the daily average concentration of PM2.5 at the city level. In an effort to alleviate the endogeneity problem, city fixed effects and temporal dummies are controlled throughout the regression work. Moreover, considering the potential confounding factors, a series of weather covariates and demographic characteristics are also involved in the estimation.

In our preferred specifications, the estimates reveal that a one standard deviation of PM2.5 (around 30 $\mu g/m^3$) increases the likelihood of inconsistency by 2.25%. The estimates prove to be remarkably significant and robust under a variety of tests. Provided that the effect size could differ under the heterogeneous demographic groups, we further replicate the regressions in sub-samples of different gender, age, education, income and jobs. The results suggest that air pollution appears to have larger effects on those young groups with high education but low income, especially for students, and the number is 5.36%.

In addition, thermal environment is also a significant factor that affects people's attention ability to accomplish a certain task. Many recent studies (e.g. Choi and Chun (2009); Lee et al. (2012)) have shown that temperature controlled at 20 °C to 25 °C is conductive to exerting the better attention ability. The similar result is found in our work. We create seven bins of average temperature in the regression, with 5 °C as each of bin width. Compared to the temperature under 5 °C, temperature between 20 °C to 25 °C decreases the inconsistency by 26.46%. The additional study on temperature could also be seen as a supplement to confirm the validity of our construction of the inconsistent index.

Focus on the attention loss due to air pollution is interesting. However, we are careful not to over-interpret the outcomes. Measuring the effect of attention

impairment falls outside the scope of this article. The results are instructive from two aspects. Firstly, as far as the survey itself is concerned, inattentive responding impacts the accuracy and validity of the data, reducing the statistical power. Secondly, the results could also be extended to other learning and work scenarios, providing a mechanism to support the study that explores the association between air pollution and labour productivity (Graff Zivin & Neidell, 2012; Liu et al., 2021), worker performance (Chang et al., 2019), school-based grade point (Balakrishnan & Tsaneva, 2021; Grineski et al., 2016), and traffic accidents (Sager, 2019; Wan et al., 2020).

The remainder of the paper is organized as follows. Section 2 describes the data sources. Section 3 explains the empirical methodology involved in the regression. Section 4 presents the main results. Section 5 further discusses the heterogeneous effects. Section 6 conducts the robust tests. Section 7 examines the placebo tests. Section 8 concludes the study.

2. Data

The objective of this paper is to investigate the potential link between air pollution and the impaired attention. Thus, the measurement of people's inattention has become the most important challenge at the beginning. Here, we provide a new thought to study the inattentive (or careless) responding existing in the online surveys, and examine whether the inattentive behaviours of the participants are affected by their exposure to air pollution.

2.1. Online survey

The online survey we mainly use for this analysis is the 'Netizen Ideology Survey' designed and implemented by Dr.Ma at the Renmin University in Beijing, that was published on the platform of the Chinese General Social Survey (CGSS). The project aims to explore the ideology of netizens from multiple perspectives, such as social identity, patriotism, nationalism, ideological stance, political and social trust, democratic values, etc.

The 'Netizen Ideology Survey' had carried out three rounds of online questionnaires in 2014, 2015 and 2017, respectively. Specifically, during May to July in 2014, July to August in 2015, and April in 2017, online questionnaires were released on the mainstream media platforms in China like Aidiaoyan (www.idiaoyan.com), Sina Weibo (www.weibo.com), Kaidi club (club.kdnet.net), and Tianya Forum (www.tianya.cn). The netizens responded to the survey voluntarily. Each round of the survey contains around fifty questions including demographic information, media use, psychological traits, social attitudes and political attitudes. Meanwhile, the survey also records the respondent's IP address, as well as the start time and end time of filling in the questionnaire online, so that we can track the location of the respondent and the date of the answering day.

There are total of 10107 participants involved in the dataset, of which 62.4% are male. The respondents are relatively young groups (54.7% are aged under 40), with high level of education (72.3% hold degrees of or higher than Bachelor) and middle-

level household income (50.6% have household income of 60k to 200k RMB)[3]. The occupational groups that account for the largest proportion are company employees (26.5%) and students (17.5%). Therefore, demographic information in our sample is also in line with the characteristics of Internet users in the contemporary media environment.

2.2. Measuring inattention

There are several methods for identifying the inattentive (or careless) respondents via the specific indicators obtained from their responses. Meade and Craig (2012) discussed the usage of five indices in their paper: (1) special items designed to have obvious right answers; (2) the inconsistent index from the pair of similar questions; (3) multivariate outliers detection; (4) time to fill out the questionnaire; (5) self-reported authenticity in answering questions. Maniaci and Rogge (2014) mainly introduced the infrequency scales and inconsistency scales to examine the inattentive responding, which are similar to the first and second indices listed above. The infrequency scales helps to detect the infrequent responding among highly skewed responses (e.g., 'I have been to every country in the world', that is almost unlikely to happen for respondents.). The inconsistency scales are built based on the inconsistent responses when people answer the similar questions.

2.3. Inconsistent index

Taking into account the characteristics of the existing surveys, we construct the inconsistent index based on the inconsistent responses from a series of identical questions presented in different ways. Among the three waves of the Netizen Ideology Survey, such questions are not always consistent within the year of 2014, 2015, and 2017.

We focus on the most representative pair of questions designed in both of 2015 and 2017 waves, that is 'Our nation is insecure and chaotic, and there could be major disturbances at any time'. and 'In spite of the challenges we are facing, our nation is still orderly and thriving as a whole'. Both questions are designed to reflect the sense of social security, with five points Likert scale ranging from strongly disagree, disagree, neutral, agree to strongly agree. As for that in 2014, the selected pair of questions are related to national satisfaction, that is 'Do you trust our national leadership? (definitely don't trust, rather don't trust, rather trust, and definitely trust)' and 'What is your attitude at the future and destiny of our country? (very optimistic, rather optimistic, not optimistic, definitely optimistic)'.

We create the binary inconsistent index with the number equal to 1 representing that the participant holds different views on the same issue. To avoid over-evaluating the inconsistency, we do not distinguish the strength of the opinions, for example, strongly disagree and disagree are classified as the same attitude, and the thing is also the same to four-item scales of trust and optimism.[4] In Table 1, it can be seen that, among 9474 observations, 28.1% of the responses are detected to be inattentive responding in our sample.

2.4. Air pollution

We collect the air pollution data from the Chinese Ministry of Environmental Protection (MEP) which is responsible for disclosing the hourly concentration of six air pollutants from the monitoring sites at each city, including PM10 ($\mu g/m^3$), PM2.5 ($\mu g/m^3$), CO (mg/m^3), NO2 ($\mu g/m^3$), and O3 ($\mu g/m^3$). The disclosure of official air quality data in China has achieved a significant improvement since 2014. On the one hand, the data are pushed forward to cover more cities. By the end of 2014, a total of 189 cities had released their real-time ambient data, and the number increased to 366 in the year of 2015 and 2017.[5] On the other hand, the accuracy of the data has been greatly enhanced. Stoerk (2016) tested Chinese data against US Embassy data for consistency with Benford's Law, and concluded that misreporting had likely ended after 2013.

It is our main interest to explore the possible link between air pollution and the inattentive behaviour of the participants in online surveys. Due to the disclosed IP address and responding time, air pollution data can be linked to the city where each user is located, as well as the current date (even hour).[6] It is worth pointing out that there could exist the missing data of several remote cities for the lack of monitors in the early years, we replace with the data from the nearest neighbouring city whose data are available. Provided that the short-term effect of ambient pollutants on brain health could sustain for at least 24 hours, our air pollution data are integrated into the daily average at the city level, which is also consistent with the previous study like Chang et al. (2019).

Among those six pollutants reported, we mainly focus on PM2.5, since the fine particle has a larger detrimental effect on brain function (Elder et al., 2006). Now our topic has been specific to explore the effect of PM2.5 on the inconsistent index we built based on the online survey. Other pollutants will also be included to conduct the joint estimation as the robustness check. The standard deviation of PM2.5 in Table 1 is reported to be 29.9 $\mu g/m^3$, which will be used as our estimation unit when we describe the effect size.

2.5. Weather

Weather is an obviously important factor in affecting the ability of sustained attention. In particular, the brainwork has been well documented to be more vulnerable in heat and cold scenarios (Yeganeh et al., 2018). Besides that, weather covariates are also a series of confounding parameters that affect the formation of air pollutants through the complex photochemical reactions, which could lead to the endogeneity of air pollution variable and the biased estimates if ignored in the regression. In order to alleviate the potential endogeneity, we attempt to include the comprehensive meteorological characteristics.

The meteorological data are collected from the National Oceanic and Atmospheric Administration (NOAA). Corresponding to the air pollutant, the hourly data are merged into daily average at the city level and connected to the inconsistent index of each respondent. The involved parameters include average temperature (°C), humidity (%), precipitation (mm), wind speed (km/h), sea-level pressure (hPa) and cloud coverage. Table 1 lists the summary statistics with the mean and standard deviation for each meteorological factor.

Table 1. Summary statistics.

Variables	Obs (1)	Mean (2)	Std. Dev. (3)	Min (4)	Max (5)
Inconsistent Index	9474	0.2814	0.4497	0	1
Air Pollutants					
PM2.5 ($\mu g/m^3$)	9476	48.8185	29.9028	0	264
NO2 ($\mu g/m^3$)	9455	36.2873	15.2255	0	105
SO2 ($\mu g/m^3$)	9486	16.1524	12.3992	0	175
CO (mg/m^3)	9485	0.9039	0.3349	0	3.5
O3_8h ($\mu g/m^3$)	8888	116.7142	48.2858	0	294
Meteorological Factors					
Avg Temperature (˚C)	9807	23.9972	5.4968	−3.9	37.6
Humidity (%)	9837	62.3949	20.4833	7.9	97.6
Precipitation (*mm*)	9130	4.2551	11.9092	0	201
Wind Speed (*km / h*)	9807	8.203	3.7086	1	48
Sea-level Pressure (*hPa*)	9464	1007.149	4.6226	992	1027.4
Cloud Coverage	9673	5.5669	2.1655	0	8

Notes: This table lists the summary statistics of the main variables involved in the paper.
Source: estimated and made by authors.

3. Methodology

In this paper, we attempt to explore the potential link between air pollution and people's inattentive behaviour. Our analysis is placed in the scenario of web-based online surveys where the inconsistent index is constructed to proxy the inattentive and careless responses of the participants. We perform the estimation by regressing the inconsistent index on air pollution, especially PM2.5, as well as a series of controls.[7]

3.1. Probit model

In this paper, the probit model is mainly adopted to estimate the probability of inattentive behaviour in online surveys due to air pollution, which is one of the most common methods to deal with the discrete choices regression. We set up the model in the following form:

$$y_{ict} = \alpha + \beta \cdot PM2.5_{ct} + \zeta \cdot CP_{ct} + \gamma \cdot W_{ct} + \delta \cdot D_i + \lambda_c + \nu_t + \epsilon_{ict} \tag{1}$$

where the inconsistent index y_{ict} is the binary variable we construct, with 1 representing the existence of the inattentive responses for the respondent i in the city c at the date t, and 0 being normal. CP_{ct} stands for a variety of multi-pollutants such as CO, SO2, NO2, and O3. W_{ct} denotes a set of weather covariates including temperature, humidity, precipitation, wind speed, sea-level pressure and cloud coverage. D_i controls the respondent i's demographic characteristics, like gender, age, education, income and jobs (Table 2).

In addition, we also incorporate the city fixed effects λ_c in the regression, so as to control for the time-invariant characteristics across cities. ν_t represents the temporal controls, including year by month fixed effects and weekday dummies, which help to control for the seasonal patterns and daily changes across a week. ϵ_{ict} is the error term. β is our main parameter of interest. The sign of the estimate is expected to be positive, indicating the effect of PM2.5 on respondents' inattentive behaviours in online surveys.

Table 2. Demographic characteristics.

		Freq. (1)	Percent (2)
Gender	Male	6308	62.4
	Female	3799	37.6
Age	18-24	1704	16.9
	25-29	1663	16.5
	30-34	1265	12.5
	35-39	885	8.8
	40-44	837	8.3
	45-49	645	6.4
	50-54	399	3.9
	55-59	140	1.4
	> 60	167	1.7
Education	No formal education	10	0.1
	Primary education	26	0.3
	Lower secondary	210	2.1
	Higher secondary	861	8.5
	Junior College	1693	16.7
	Bachelor	5323	52.7
	Master	1478	14.6
	Doctorate	506	5.0
Household Income	< 5,000	319	3.2
	5,000 − 10,000	329	3.3
	10,000 − 20,000	497	4.9
	20,000 − 40,000	785	7.8
	40,000 − 60,000	1308	12.9
	60,000 − 100,000	2266	22.4
	100,000 − 200,000	2852	28.2
	200,000 − 500,000	1341	13.3
	500,000 − 1,000,000	279	2.8
	1,000,000 − 2,000,000	72	0.7
	> 2,000,000	59	0.6
Job	Worker(Agriculture, Forestry, Animal husbandry and Fishery)	136	1.3
	Factory worker	715	7.1
	Senior government officials	323	3.2
	Middle or senior manager in state-owned enterprises	300	3.0
	Judiciary	476	4.7
	Soldier	28	0.3
	Employee in company	2680	26.5
	Self-employed	483	4.8
	Middle or senior managers in private enterprises	538	5.3
	Freelancer	483	4.8
	Student	1773	17.5
	Healthcare worker	223	2.2
	Teacher (Primary and secondary school)	424	4.2
	College, university or research institute	586	5.8
	News media	120	1.2
	Lawyer	77	0.8
	Arts and cultural	47	0.5
	Non-governmental organization (NGO)	31	0.3
	Mass organizations	33	0.3
	Unemployed	209	2.1
	Others	422	4.2

Notes: This table displays the frequency and percentage of demographic characteristics analysed in the study.
Source: estimated and made by authors.

3.2. Nonlinear effects

In the probit model discussed above, air pollution is involved in the linear form of daily average concentration at each city. However, people's response to different levels of ambient pollutants could be uneven and more likely to increase at a diminishing rate due to the potential adaptability to the dirty air. With that in mind, we

categorize the concentration of PM2.5 into five groups: $(0, 35]$ (the reference group), $(35, 75]$, $(75, 115]$, $(115, 150]$, and $(150, 250]$ according to the Technical Regulation on Ambient Air Quality Index (on trial) (HJ 633-2012). Each group corresponds to the different levels of ambient pollution, and is converted from the relevant Air Quality Index (AQI).[8] The five bins of PM2.5 will be introduced into the regression and replace the daily average. After transformed into the average marginal effect, each estimate could be interpreted as how much increase in the probability of inconsistent response when the participants are exposed to higher pollution levels than in clean weather with PM2.5 under $35\mu g/m^3$.

4. Results

4.1. Effect of PM2.5 on inattentive behaviour

Table 3 presents the effect of PM2.5 on inattentive behaviour of the participants that is proxied by the inconsistent index. All the regressions are estimated by probit model introduced above, and the results reported have been transformed into the average marginal effects in the form of percentage. From left to right, additional controls are added to complete the model structure. Moreover, considering the existence of the heteroscedastic and autocorrelated problems in error terms, we also cluster the robust standard errors at the city level. It can be seen that all the estimates reveal the positive correlation between PM2.5 and inconsistent index, most of which are significant at 1% level.

In column (1), city fixed effects are included in the basic regression and help to control for the unobserved time-invariant characteristics of each city, which might affect the citizensbehaviour and local pollution' behaviour and local pollution level. In column (2), year by month fixed effects and weekday fixed effects are considered to control for the seasonal and periodic variations that are constant across individuals. As expected, time fixed effects play an important role in people's attention ability.

Table 3. The effect of PM2.5 on inconsistent responses.

Dependent variable: Inconsistent Index	City fixed effects (1)	Temporal controls (2)	Weather controls (3)	Demographic characteristics (4)	Multi-pollutants (5)
PM2.5	0.0383**	0.0675***	0.0667***	0.0596***	0.0751***
	(0.0187)	(0.0178)	(0.0184)	(0.0186)	(0.0246)
Additional Controls					
City FEs	Y	Y	Y	Y	Y
Year by month FEs	N	Y	Y	Y	Y
Weekday FEs	N	Y	Y	Y	Y
Weather covariates	N	N	Y	Y	Y
Demographics	N	N	N	Y	Y
Co-pollutants	N	N	N	N	Y
Observations	8094	8094	7462	7460	6974

Notes: This table presents the effect of PM2.5 on peoples' inconsistent responses existing in the web-based survey. Additional controls are added from left to right. All the estimates have been adjusted into the marginal effects in the form of percentage. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. Co-pollutants include NO_2, SO_2, CO, and O_3. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%). Source: estimated and made by authors.

The outcome increases a bit in magnitude. Weather covariates in column (3) contain average temperature, average humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics in column (4) involve gender, age, education, household income, and jobs. From column (2) to column (4), the outcomes are quite robust in both magnitude and significance. Column (5) summarizes our preferred estimation under full controls with the inclusion of multi-pollutants (NO2, SO2, CO, and O3). The result can be explained that a one standard deviation increase in daily average PM2.5 will increase the probability of being inattentive by 2.25%.[9]

4.2. Nonlinear effect of PM2.5 on inattentive behaviour

In order to examine the nonlinear effect of air pollution on inattentive behaviours, we construct five bins for PM2.5 (with 40 $\mu g/m^3$ as bin width) according to the guidance of the Technical Regulation HJ 633-2012, and implement the identical regressions as those for linear estimation. As Table 4 shows, the same controls are added from left to right, and the estimates appear to be rather robust especially when models have enough controls. The most striking result to emerge from the outcomes is that the marginal effects have larger values in magnitude when exposed to a higher level of PM2.5 compared to the reference group with PM2.5 under 35 $\mu g/m^3$, implying that people are more likely to be inattentive as air pollution worsens.

Table 4. The non-linear effects of PM2.5 on inconsistent responses.

Dependent variable: Inconsistent Index	City fixed effects (1)	Temporal controls (2)	Weather controls (3)	Demographic characteristics (4)	Multi-pollutants (5)
Reference group: PM2.5 ≤ 35 ($\mu g/m^3$)					
PM2.5 ⊂ (35, 75]	2.6564*	3.242**	2.5981*	2.1746	2.9715*
	(1.5855)	(1.4635)	(1.5511)	(1.4719)	(1.7513)
PM2.5 ⊂ (75, 115]	4.7325**	5.9138***	5.6080***	5.0068***	6.2088***
	(2.0484)	(1.5626)	(1.7247)	(1.7076)	(2.2184)
PM2.5 ⊂ (115, 150]	3.466	6.5064**	6.2867**	5.6464*	7.4188*
	(2.8649)	(3.0951)	(3.2073)	(3.1116)	(4.0301)
PM2.5 ⊂ (150, 250]	3.8087	8.4081*	9.3364*	9.4531*	9.3921
	(4.5965)	(5.0032)	(5.3241)	(5.3546)	(6.5939)
Additional Controls					
City FEs	Y	Y	Y	Y	Y
Year by month FEs	N	Y	Y	Y	Y
Weekday FEs	N	Y	Y	Y	Y
Weather covariates	N	N	Y	Y	Y
Demographics	N	N	N	Y	Y
Co-pollutants	N	N	N	N	Y
Observations	8094	8094	7462	7460	6972

Notes: This table presents the non-linear effects of PM2.5 on peoples' inconsistent responses existing in the web-based survey. PM2.5 is categorized into five groups (including the omitted group) according to Technical Regulation on Ambient Air Quality Index (on trial) (HJ 633-2012). PM2.5 equal or less than 35 $\mu g/m^3$ is the reference group for all the regressions. Additional controls are added from left to right. All the estimates have been adjusted into the marginal effects in the form of percentage. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. Co-pollutants include NO2, SO2, CO, and O3. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%).
Source: estimated and made by authors.

Table 5. Air quality index.

Level (1)	AQI (2)	PM2.5 ($\mu g/m^3$) (3)	Health Messages (4)
Excellent	0-50	0-35	No air pollution problem.
Good	51-100	36-75	Air quality is considered acceptable.
Light Polluted	101-150	76-115	Air pollution affets the sentive group, like children, the elderly, and the patients with respiratory disease or heart disease.
Moderately Polluted	151-200	116-150	All people will be seriously affected, and preventive measures should be taken when necessary.
Heavily Polluted	201-300	151-250	
Severely Polluted	301-500	251-500	

Notes: This table lists the scale of the Air Quality Index (AQI) designed by the Technical Regulation on Ambient Air Quality Index (on trial) (HJ 633-2012). AQI is categorized into six levels: 0-50, 51-100, 101-150, 151-200, 201-300, and greater than 300, ranging from excellent to severely polluted. The corresponding daily concentration of PM2.5 ($\mu g/m^3$) has been converted and reported in column (3). The health-based guidelines are also displayed in column (4). Source: estimated and made by authors.
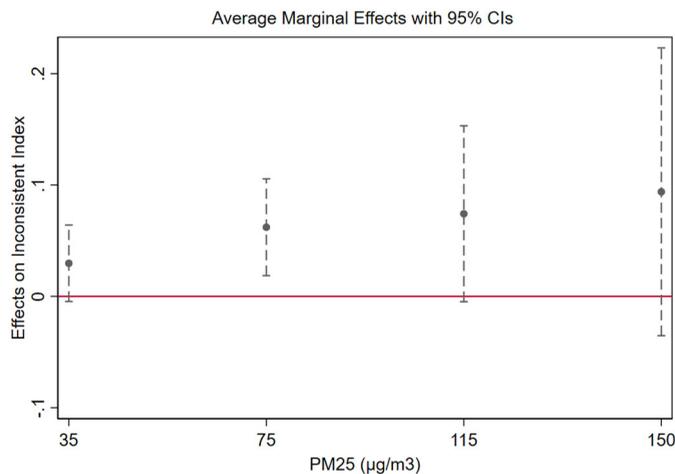
Figure 1. The non-linear effects of PM2.5 on inconsistent index under full controls.
Notes: This graph displays the non-linear impacts of PM2.5 on people's inconsistent responses compared to the reference group under full controls. PM2.5 is categorized into five groups: PM2.5 less than $35\mu g/m^3$ (reference group), PM2.5 $\subset (35, 75]$, PM2.5 $\subset (75, 115]$, PM2.5 $\subset (115, 150]$, and PM2.5 $\subset (150, 250]$. The figure plots the estimates and their 95% confidence intervals according to the coefficients shown in column (5) of Table 4. Source: estimated and made by authors.

To be more clear, Figure 1 plots the estimates and 95% confidence intervals according to the coefficients under full controls shown in column (5) of Table 4. Interestingly, people's response to air quality increases quickly when exposed to low-to-medium levels of air pollution. According to the AQI categories shown in Table 5, a 'Good' day with PM2.5 between 36 and 75 $\mu g/m^3$ leads to 2.97% increase in inconsistent responses. Moreover, a 'Light Polluted' day with PM2.5 between 76 and 115 $\mu g/m^3$ leads to 6.21% increase in inconsistent responses. Until now, an approximate linear growth trend can be found when PM2.5 is less than 115 $\mu g/m^3$. Subsequently, the growth rate of people's response to air pollution begins to stabilize for their increased adaptability to dirty air, which is also consistent with the previous studies (Graff Zivin & Neidell, 2012; He et al., 2019; Qin & Zhu, 2018). A 'Moderately

Polluted' day with PM2.5 ranging from 116 to 150 $\mu g/m^3$ causes a growth of 7.42%, not much increase compared with the previous group. Notably, the curve also becomes flat between those two intervals.

Taken together, our results suggest a positive link between air pollution and people's inattentive behaviours, as well as a diminishing increase rate of the effects under high-level polluted days with PM2.5 larger than 115 $\mu g/m^3$. Since 96.76% of the days in our sample have PM2.5 less than 115 $\mu g/m^3$, we will continue to estimate the effect of PM2.5 in linear form during the following analysis.

## 4.3. Effect of temperature on inattentive behaviour

Temperature is an essential factor that affects people's brain function (Yeganeh et al., 2018). It has been demonstrated that hot (32.22°C or above) and cold (10°C or less) temperature decreases the performance on a wide range of mental work like attention, perception and memory (Pilcher et al., 2002). Therefore, such results should be verified again in our research. Here, the analysis on temperature can also be seen as a supplement to confirm the validity of our construction of the inconsistent index. We categorized daily average temperature into seven bins (with 5°C as each bin width): less than 5°C (the reference group), [5, 10), [10, 15), [15, 20), [20, 25), [25, 30), and above 30°C. Table 6 presents the marginal effects of different temperatures on people's inattentive behaviour, compared to that in the cold scenario under 5°C.

Table 6. The non-linear effects of temperature on inconsistent responses.

Dependent variable: Inconsistent Index	City fixed effects (1)	Temporal controls (2)	Demographic characteristics (3)	Pollutant (4)
Reference group: T < 5 (°C)				
T ⊂ [5, 10)	−29.195***	−28.7514***	−27.5505***	−23.8084**
	(11.8038)	(11.2526)	(10.6239)	(10.0638)
T ⊂ [10, 15)	−26.0561**	−26.4272**	−26.0148**	−21.753**
	(12.0027)	(11.4924)	(10.8143)	(10.2034)
T ⊂ [15, 20)	−31.3169***	−30.6995***	−29.5123***	−25.9037***
	(11.5783)	(11.1016)	(10.4689)	(9.8259)
T ⊂ [20, 25)	−36.5537***	−31.2391***	−29.5025***	−26.4647***
	(11.6307)	(11.2647)	(10.6175)	(10.0505)
T ⊂ [25, 30)	−31.8662***	−28.6345***	−27.2701***	−25.3831***
	(11.5895)	(11.3821)	(10.7356)	(10.1898)
T ≥30	−30.5227***	−27.3463**	−25.9978**	−23.566**
	(11.8683)	(11.5572)	(10.9586)	(10.3859)
Additional Controls				
City FEs	Y	Y	Y	Y
Year by month FEs	N	Y	Y	Y
Weekday FEs	N	Y	Y	Y
Weather covariates	Y	Y	Y	Y
Demographic	N	N	Y	Y
Pollutant	N	N	N	Y
Observations	7702	7702	7700	7460

Notes: This table demonstrates the non-linear effects of temperature on peoples' inconsistent responses when filling out the online survey. The average temperature is categorized into seven bins, with 5°C as each of bin width. Temperature less than 5°C is the reference group for all the regressions. Additional controls are added from left to right. All the estimates have been adjusted into the marginal effects in the form of percentage. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity, wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. PM2.5 is considered when controlling the pollutant. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%).
Source: estimated and made by authors.
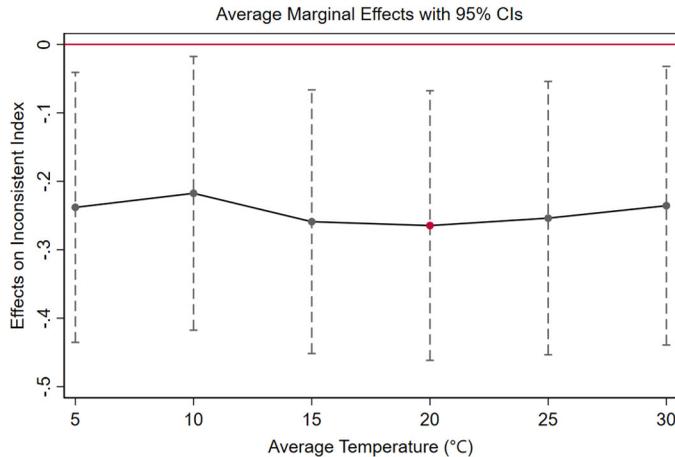
Figure 2. The non-linear effect of temperature on inconsistent responses under full controls.
Notes: This graph displays the non-linear impacts of average temperature on people's inconsistent responses compared to the reference group under full controls. The average temperature is categorized into seven groups, with 5°C as each of bin width and temperature less than 5°C as the reference group. The figure plots the estimates and their 95% confidence intervals according to the coefficients shown in column (5) of Table 6. The red dot in the figure marks the temperature that has the least impact on the inconsistent index.
Source: estimated and made by authors.

Identically, the same additional controls are added from left to right as we did in previous tables. The estimates are quite robust in both magnitude and significance.

What is striking about the figures in this table is that all the estimates are negative in relatively large value, suggesting that temperature above 5°C is much better for people to maintain sustained attention. In addition, the effects appear to be falling first and then rising. Correspondingly, Figure 2 depicts the non-linear estimates under full controls and their 95% confidence intervals according to the coefficients shown in column (5) of Table 6. The overall trend seems to be plausibly a U-shape pattern, in which the red dot around 20°C marks the temperature that has the least impact on the inconsistent index. It can be concluded that compared to the temperature under 5 $^{\circ}$C, temperature between 20 $^{\circ}$C to 25 $^{\circ}$C decreases the inattentive responses by 26.46%. The result we obtain is fairly consistent with that in the recent findings. The studies like Choi and Chun (2009) and Lee et al. (2012) have also shown that temperature controlled at 20 $^{\circ}$C to 25 $^{\circ}$C is conductive to exerting the better attention ability.

4.4. Various demographic characteristics and inattentive behaviour

Besides the external factors like air pollution and weather discussed above, the internal demographic characteristics could also affect the individual's attention ability to a large extent. Tables 7 and 8 report the impacts of various demographic variables on inconsistent responses, like gender, age, education and income. All the estimates have been transformed into the marginal effects in percentage compared to the reference group. Correspondingly, Figure 3 plots four graphs one by one according to the estimates and their 95% confidence intervals reported in the tables.

Table 7. The impact of various demographic characteristics on inconsistent responses.

| Dependent variable: Inconsistent Index	City fixed effects (1)	Temporal controls (2)	Weather covariates (3)	Pollutant (4)
Reference group: Male				
Gender	−1.0424	−1.4242	−1.7656*	−1.6227
	(1.0526)	(1.0324)	(1.0863)	(1.0976)
Reference group: Age ≤ 24				
Age ⊂ [25, 34]	4.0213***	3.6813**	4.1258**	4.4373***
	(1.5967)	(1.6205)	(1.7374)	(1.7732)
Age ⊂ [35, 49]	2.9196*	1.8993	2.5755	2.5133
	(1.7939)	(1.8712)	(2.022)	(2.0624)
Age ≥ 50	2.0908	1.7151	2.3552	2.1093
	(2.2378)	(2.3645)	(2.4926)	(2.532)
Reference group: Doctorate				
Master	4.2429	4.4676	3.9736	3.8673
	2.9771	(2.9344)	(3.0358)	(3.1385)
Bachelor	6.1958**	6.2401**	5.8092*	5.6269*
	(3.1969)	(3.1048)	(3.2143)	(3.3288)
Junior College	9.0082***	9.3579***	8.4531**	8.4342**
	(3.2962)	(3.2134)	(3.4853)	(3.5575)
Higher secondary	9.1431***	9.4012***	7.8809**	7.1898*
	(3.666)	(3.6173)	(3.7339)	(3.8184)
Lower secondary	5.9492	6.9885	6.3343	6.1583
	(4.9905)	(4.9669)	(5.2711)	(5.3719)
Primary education	20.6778*	23.9162*	22.6763*	20.8256*
	(12.3491)	(12.5396)	12.3506	(12.7306)
No formal education	37.1411	41.2522	–	–
	(25.4231)	(27.2455)	–	–
Additional Controls				
City FEs	Y	Y	Y	Y
Year by month FEs	N	Y	Y	Y
Weekday FEs	N	Y	Y	Y
Weather covariates	N	N	Y	Y
Demographic	Y	Y	Y	Y
Pollutant	N	N	N	Y
Observations	8364	8364	7700	7460

Source: estimated and made by authors.

4.4.1. Gender and age

As for gender, the female seem to be slightly better at concentration than the male, and the difference is 1.77% significant at 10% level reported in column (3) of Table 7 without controlling for pollution. Besides, the more surprising result is reflected in the performance of different age groups. People aged from 25-34 appear to be the most inattentive, with 4.44% more inconsistent responses compared to the younger group under 25. The effects are statistically significant at 1% and robust under additional controls from column (1) to column (4). Young adults from 25-34 years of age are physically young and energetic, and less likely to have brain defects. The potential reason for their unsatisfactory performance on attention could be a problem with pressure from both work and life, which should be paid more attention to. Interestingly, as we continue to look down, although not significant, the performance of attention is numerically better when age increases shown in Panel (b) of Figure 3.

4.4.2. Education

In regard to education, from the top to the bottom, different degree levels are listed from doctorate (the highest, reference group) to no formal education (the
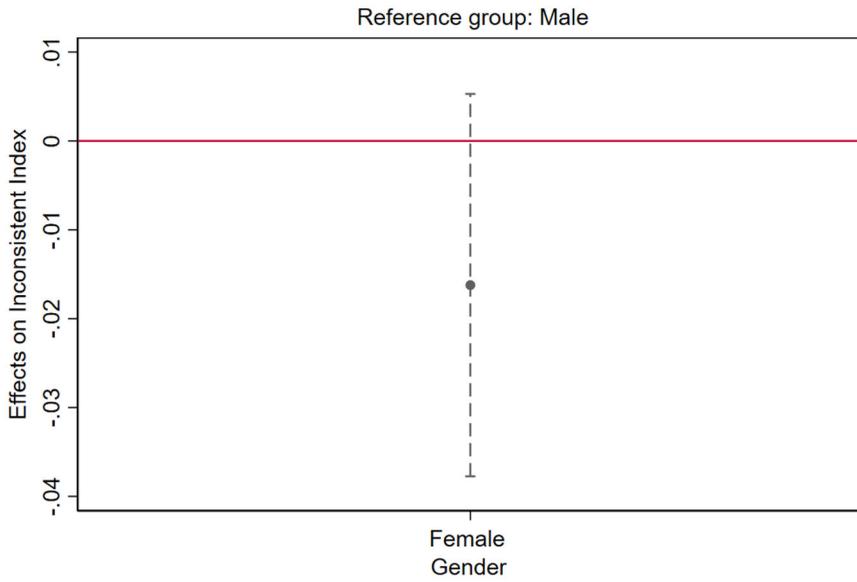
(a) The impact of gender on inconsistent responses
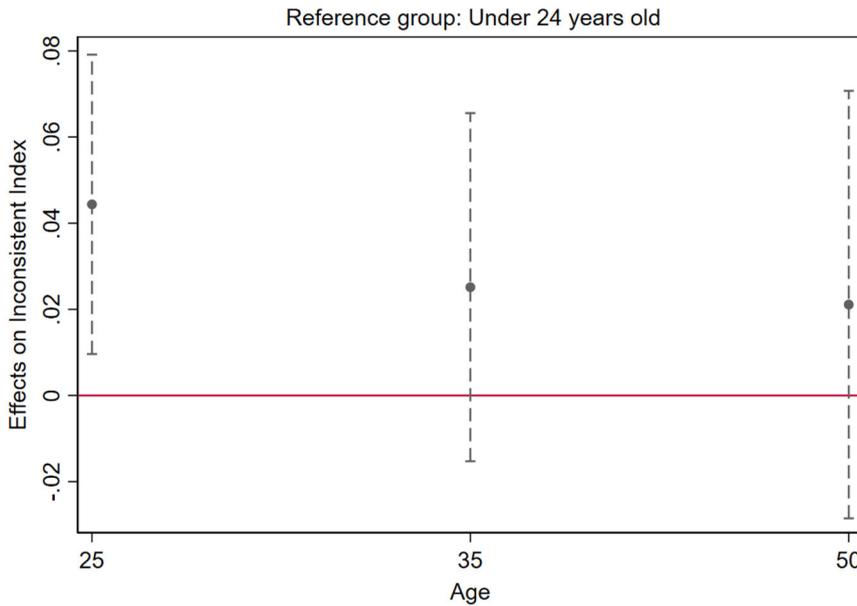
(b) The impact of different ages on inconsistent responses

Figure 3. The marginal effects of demographic characteristics on inconsistent response.
Source: estimated and made by authors.

lowest). When looking down, the estimates show a plausibly increasing trend (see Panel (c) in Figure 3), implying that the participants with lower education levels are more likely to have inattentive responses. Junior college or higher secondary
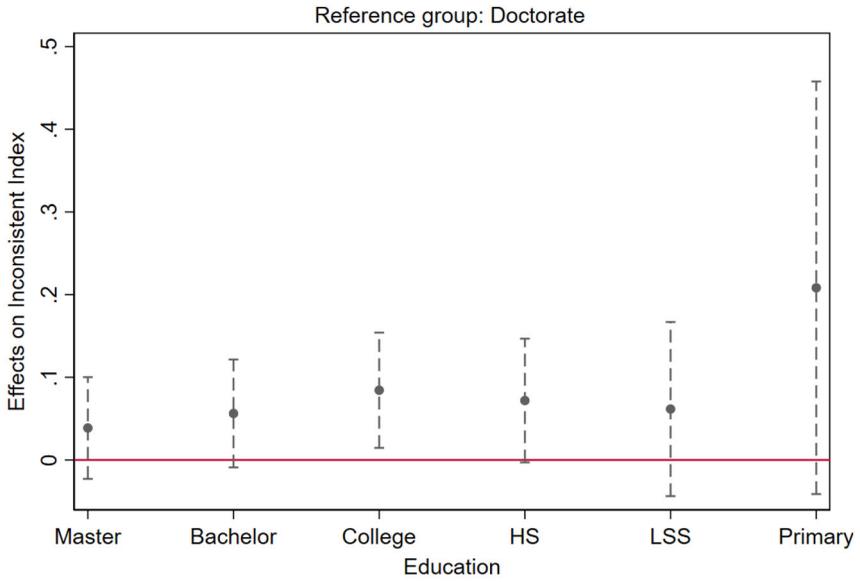
(c) The impact of different levels of education on inconsistent responses
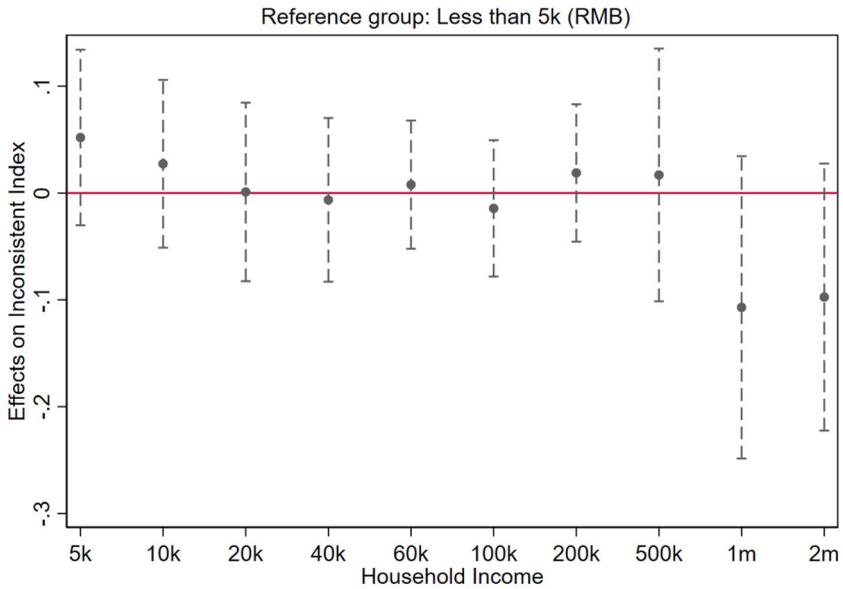
(d) The impact of different household income levels on inconsistent responses

Figure 3. Continued

groups have a 7-8% higher probability of inconsistent responses than the doctorate group, and the elementary group is even higher with the number being around 20%.

Table 8. (Continued) The impact of various demographic characteristics on inconsistent responses.

Dependent variable: Inconsistent Index	City fixed effects (1)	Temporal controls (2)	Weather covariates (3)	Pollutant (4)
Reference group: Household income < 5,000 RMB				
Income $\subset [5,000, 10,000)$	3.9067	3.3043	4.2626	5.1972
	(3.9222)	(4.0017)	(4.1539)	(4.19)
Income $\subset [10,000, 20,000)$	3.3297	2.9306	2.2251	2.7335
	(3.7413)	(3.7586)	(3.9332)	(4.0063)
Income $\subset [20,000, 40,000)$	−2.1364	−1.2052	−1.0871	0.0983
	(3.8978)	(3.8721)	(4.1946)	(4.2626)
Income $\subset [40,000, 60,000)$	−2.3517	−2.0309	−1.5876	−0.6419
	(3.7112)	(3.692)	(3.8514)	(3.9111)
Income $\subset [60,000, 100,000)$	0.0542	0.4868	−0.2547	0.7815
	(2.9118)	(2.8998)	(3.0141)	(3.0627)
Income $\subset [100,000, 200,000)$	−0.608	−0.4965	−2.0628	−1.4385
	(3.1858)	(3.1654)	(3.2132)	(3.2561)
Income $\subset [200,000, 500,000)$	2.2458	2.1208	0.9047	1.8797
	(3.1638)	(3.1944)	(3.2547)	(3.2835)
Income $\subset [500,000, 1,000,000)$	−1.2296	−0.4282	0.3965	1.69
	(5.6368)	(5.6175)	(5.978)	(6.0349)
Income $\subset [1,000,000, 2,000,000)$	−15.1545**	−13.6949**	−11.7855	−10.7054
	(7.3237)	(7.0873)	(7.3106)	(7.2181)
Income $>2,000,000$	−12.3492*	−12.0752*	−11.11*	−9.7432
	(6.8045)	(6.5851)	(6.2355)	(6.3777)
Additional Controls				
City FEs	Y	Y	Y	Y
Year by month FEs	N	Y	Y	Y
Weekday FEs	N	Y	Y	Y
Weather covariates	N	N	Y	Y
Demographic	Y	Y	Y	Y
Pollutant	N	N	N	Y
Observations	8364	8364	7700	7460

Notes: This table presents the impact of various demographic characteristics on inconsistent responses, including gender, age, education and income. All the estimates have been adjusted into the marginal effects in the form of percentage. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. PM2.5 is considered when controlling the pollutant. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%).
Source: estimated and made by authors.

4.4.3. Household earnings

Table 8 continues to present the association between income and inconsistent responses, where household earnings are categorized into 11 groups from the lower one less than 5,000 RMB to the highest one above 2,000,000 RMB. From the data, it can be seen that there is no obvious trend as household income increases. However, for the upper class whose income is above 1,000,000 RMB, people are more likely to have better attention ability, and their probability of inconsistent responses has dropped drastically with around 10% decrease compared to the low-income class.

Together these results from education and income provide important insights into the link between attention ability and work performance. Our outcome indicates that people engaged in jobs that require a high level of education or high-paying jobs are more likely to have better attention ability. From another perspective, it is attention that helps them to have outstanding work performance on cognitive-demanding tasks, such as advanced learning, and high-productivity jobs corresponding to high income.

5. Heterogeneous effects

As discussed above, different demographic characteristics of respondents may show different levels of attentiveness. In this section, we further analyse whether the influence of air pollution on people's inattentive behaviour will show heterogeneous changes under different demographic characteristics. Tables 9–13 demonstrate the marginal effects of PM2.5 on inconsistent responses under full controls among the varying gender, age, education, household income and occupational groups.

5.1. Heterogeneous effects of PM2.5 under different genders and ages

In Table 9, column (1) and column (2) conduct the regressions under female and male, respectively. Column (3) replicates the preferred result in column (5) of Table 3 based on the whole sample. It seems that PM2.5 affects men more than women, with larger values in magnitude and significance. Table 10 displays the effect of PM2.5 among groups with increasing average age. Given the limited subsample size at different ages, people in the higher group are added from left to right, so that the average age of the samples rises sequentially, in which we explore the impact of air pollution at different age levels. There is a plausibly decreasing trend in the estimates as the average age increases, suggesting that the attention ability of youth and young adult groups is most vulnerable to air pollution. A one standard deviation of PM2.5 increases the likelihood of inattentive responses by 4.05% among the youth aged of 18-24.

5.2. Heterogeneous effects of PM2.5 under different education levels

Table 11 implements the similar regressions as above with decreasing level of education added from column (1) to column (5), so that the effect of PM2.5 could be studied among the various levels of education. It can be seen that higher education

Table 9. The effects of PM2.5 on inconsistent responses under different gender groups.

Dependent variable: Inconsistent Index	Female (1)	Male (2)	All (3)
PM2.5	0.0579	0.0829**	0.0751***
	(0.0436)	(0.0368)	(0.0246)
Additional Controls			
City FEs	Y	Y	Y
Year by month FEs	Y	Y	Y
Weekday FEs	Y	Y	Y
Weather covariates	Y	Y	Y
Demographic	Y	Y	Y
Co-pollutants	Y	Y	Y
Observations	2513	4277	6974

Notes: This table reflects the impact of PM2.5 on inconsistent responses under female (in column (1)) and male groups (in column (2)), respectively. Column (3) replicates the estimates shown in column (5) of Table 3 based on the whole sample. Regressions are operated under full controls. All the estimates have been adjusted into the marginal effects in the form of percentage. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. Co-pollutants include NO2, SO2, CO, and O3. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%).
Source: estimated and made by authors.

Table 10. The effect of PM2.5 on inconsistent responses among groups with increasing average age.

Dependent variable: Inconsistent Index	Age 18-24 (1)	Add age 25-29 (2)	Add age 30-34 (3)	Add age 35-39 (4)	Add age 40-44 (5)	Add age 45-49 (6)	Add age 50-54 (7)	Add age 55-59 (8)	All add 60+ (9)
PM2.5	0.1356**	0.1051**	0.1158***	0.0955***	0.0891***	0.0842***	0.0741***	0.0747***	0.0751***
	(0.0668)	(0.0445)	(0.0445)	(0.0352)	(0.0288)	(0.0261)	(0.0265)	(0.0249)	(0.0246)
Additional Controls									
City FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year by month FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weekday FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather covariates	Y	Y	Y	Y	Y	Y	Y	Y	Y
Demographic	Y	Y	Y	Y	Y	Y	Y	Y	Y
Co-pollutants	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1390	2881	4064	4927	5727	6345	6688	6818	6974

Notes: This table presents the effect of PM2.5 on inconsistent responses among groups with increasing average ages. Regressions are operated under full controls. All the estimates have been adjusted into the marginal effects in the form of percentage. From column (1) to column (9), more senior age groups are added in the analysis, so that the average age of each group gradually increases. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. Co-pollutants include NO2, SO2, CO, and O3. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%).
Source: estimated and made by authors.

Table 11. The impact of PM2.5 on inconsistent responses among groups with decreasing average education levels.

Dependent variable: Inconsistent Index	Bachelor, Master and Doctorate (1)	Add Junior College (2)	Add Secondary level (4)	All (5)
PM2.5	0.1157***	0.0836***	0.0738***	0.0751***
	(0.0317)	(0.0268)	(0.0250)	(0.0246)
Additional Controls				
City FEs	Y	Y	Y	Y
Year by month FEs	Y	Y	Y	Y
Weekday FEs	Y	Y	Y	Y
Weather covariates	Y	Y	Y	Y
Demographic	Y	Y	Y	Y
Co-pollutants	Y	Y	Y	Y
Observations	5043	6241	6960	6974

Notes: This table presents the impact of PM2.5 on inconsistent responses among groups with decreasing average education levels. Regressions are operated under full controls. All the estimates have been adjusted into the marginal effects in the form of percentage. From column (1) to column (5), samples with decreasing levels of education are added to the analysis. The average education level gradually decreases from left to right. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. Co-pollutants include NO2, SO2, CO, and O3. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%).
Source: estimated and made by authors.

groups are more sensitive to air quality although they are less likely to have inattentive responses that have been discussed in the previous section. The estimate in column (1) indicates that a one standard deviation in PM2.5 increases the probability of inattentive responses by 3.46% among Bachelor, Master and Doctorate groups.

Table 12. The impact of PM2.5 on inconsistent responses among groups with increasing average household income.

Dependent variable: Inconsistent Index	Low income (1)	Add Middle class (2)	Add Upper middle class (4)	All (5)
PM2.5	0.2044**	0.0723**	0.0757***	0.0751***
	(0.0919)	(0.0365)	(0.0253)	(0.0246)
Additional Controls				
City FEs	Y	Y	Y	Y
Year by month FEs	Y	Y	Y	Y
Weekday FEs	Y	Y	Y	Y
Weather covariates	Y	Y	Y	Y
Demographic	Y	Y	Y	Y
Co-pollutants	Y	Y	Y	Y
Observations	644	6241	6685	6974

Notes: This table presents the effect of PM2.5 on inconsistent responses among groups with increasing average household income. Regressions are operated under full controls. All the estimates have been adjusted into the marginal effects in the form of percentage. From column (1) to column (5), higher-income groups are added in the analysis, so that the average household income gradually increases. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. Co-pollutants include NO2, SO2, CO, and O3. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%).
Source: estimated and made by authors.

5.3. Heterogeneous effects of PM2.5 under different income levels

The impacts of PM2.5 among different income groups are highlighted in Table 12. Column (1) runs the regression under the low-income group whose household earnings are less than 20,000. It can be seen that air pollution has the most significant impact on the lower-income people, that is a one standard deviation in PM2.5 increases the probability of inattentive responses by 6.11%, about three times the average. The middle class added in column (2) refers to the group between 20,000 to 100,000, and the upper-middle class in column (3) refers to be between 100,000 to 500,000, among which there is no significant difference found.

5.4. Heterogeneous effects of PM2.5 under different jobs

In Table 13, we select four occupational subsamples (worker, employee in company, official and manager, and student) that have relatively large sizes, and conduct the individual regressions.[10] Interestingly, students are observed to be the most vulnerable, among which the elevated air pollution in one standard deviation increases 5.36% probability of inattentive responses. Official and manager are also quite sensitive, with the number of 4.33%.

6. Robust tests

Table 14 presents the alternative models to estimate the effect of air pollution on inattentive behaviours. Column (1) adopts the probit model in our main analysis and replicates the preferred outcome listed in column (5) of Table 3. Column (2) and column (3) re-run the identical regressions under the same full controls under the linear model and the logit model, respectively. The estimate in column (3) has also been

Table 13. The impact of PM2.5 on inconsistent responses among different occupational groups.

Dependent variable: Inconsistent Index	Worker (1)	Employee in company (2)	Official and manager (3)	Student (4)	All (5)
PM2.5	0.0391	−0.0272	0.1447*	0.1792**	0.0751***
	(0.1097)	(0.0750)	(0.0839)	(0.0796)	(0.0246)
Additional Controls					
City FEs	Y	Y	Y	Y	Y
Year by month FEs	Y	Y	Y	Y	Y
Weekday FEs	Y	Y	Y	Y	Y
Weather covariates	Y	Y	Y	Y	Y
Demographic	Y	Y	Y	Y	Y
Co-pollutants	Y	Y	Y	Y	Y
Observations	644	1843	665	1023	6974

Notes: This table presents the effect of PM2.5 on inconsistent responses among different occupational groups, such as worker (agriculture, forestry, animal husbandry, Fishery, and factory), employee in company, official manager, and student. Regressions are operated under full controls. All the estimates have been adjusted into the marginal effects in the form of percentage. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. Co-pollutants include NO2, SO2, CO, and O3. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%). Source: estimated and made by authors.

Table 14. Alternative models to estimate the effect of air pollution on inconsistent responses.

Dependent variable: Inconsistent Index	Probit Model (1)	Linear Model (2)	Logit Model (3)
PM2.5	0.0751***	0.0705***	0.0751***
	(0.0246)	(0.0247)	(0.0245)
Additional Controls			
City FEs	Y	Y	Y
Year by month FEs	Y	Y	Y
Weekday FEs	Y	Y	Y
Demographic	Y	Y	Y
Co-pollutants	Y	Y	Y
Observations	6974	7141	6974

Notes: This table presents the alternative models to estimate the effect of PM2.5 on inconsistent responses. Column (1) replicates the preferred outcome listed in column (5) of Table 3. Column (2) and column (3) re-run the identical regressions under the same full controls via the linear model and the logit model, respectively. The estimate in column (3) has also been adjusted into the marginal effects in the form of percentage. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. Co-pollutants include NO2, SO2, CO, and O3. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%). Source: estimated and made by authors.

adjusted into the marginal effects in the form of percentage, so that all the estimates now can be comparable in values. We find that the estimates remain quite robust across columns in both magnitude and significance.

Recall that the inattentive behaviours in our study are measured by the inconsistent responses existing in online surveys. In Table 15, we follow the introduction in Meade and Craig (2012) and construct the alternative indices based on the characteristics of our existing sample. Column (1) reports our preferred way in which inconsistent index is used as a proxy for the inattentive or careless response, and the estimate reprints the outcome listed in column (5) of Table 3. Column (2) and column (3) construct the dummy index (with 1 being inattentive and o being normal) based on people's responding time, where fill-in time exceeding one (or half) standard

Table 15. Alternative ways to measure the inattentive behaviours.

	Inconsistent index Preferred (1)	Responding time +/- 1 Std Dev (2)	+/- 1/2 Std Dv (3)	Self-report authenticity Discrete (4)	Binary (5)
PM2.5	0.0751***	0.0371***	0.0514***	−0.0018***	0.0494*
	(0.0246)	(0.0186)	(0.0152)	(0.00058)	(0.027)
Observation	6974	8264	8264	5246	5246

Notes: This table presents the alternative ways to measure the inattentive behaviours. Column (1) reports our preferred way in which the inconsistent index is used as a proxy for the inattentive or careless response, and replicates the preferred outcome listed in column (5) of Table 3. Column (2) and column (3) construct the dummy index (with 1 being inattentive) based on people's responding time, where fill-in time exceeding one (or half) standard deviation above or below the mean is considered to be inattentive behaviour. Column (4) and column (5) implement the estimations based on the self-report authenticity, that is either in discrete form under linear model or binary variable (with 1 being not quite true) under probit model.
Source: estimated and made by authors.

Table 16. Placebo tests.

Dependent variable: Inconsistent Index	Preferred (1)	Reverse date Order (2)	Reverse alphabetical order by city (3)
PM2.5	0.0751***	0.0113	−0.0240
	(0.0246)	(0.0193)	(0.0219)
Additional Controls			
City FEs	Y	Y	Y
Year by month FEs	Y	Y	Y
Weekday FEs	Y	Y	Y
Demographic	Y	Y	Y
Co-pollutants	Y	Y	Y
Observations	6974	6747	6705

Notes: This table reports the placebo tests by replacing air pollution in our sample with other disparate data. Column (1) replicates the preferred outcome listed in column (5) of Table 3. Column (2) and column (3) re-construct the regressions based on air pollutants in other dates (reverse date order in the sample) and in other cities (reverse-alphabetic order in the sample), respectively. All the estimates have been adjusted into the marginal effects in the form of percentage. Temporal controls include year by month fixed effects and weekday dummies. Weather covariates contain average temperature, humidity (and their interaction term), wind speed, cloud coverage, sea-level pressure, and precipitation. Demographic characteristics involve gender, age, education, household income, and jobs. Co-pollutants include NO2, SO2, CO, and O3. Robust standard errors reported in parentheses are clustered at the city level. Asterisk denotes the significance level (* significant at 10%, ** significant at 5%, *** significant at 1%).
Source: estimated and made by authors.

deviation above or below the mean is considered to be inattentive behaviour. Both estimates have been adjusted into the marginal effects in the form of percentage.

Besides that, the last two columns further implement the estimations based on the self-report authenticity, which is measured on a scale of 1 (completely untrue) to 10 (completely true) in our sample. Column (4) incorporates the discrete self-report variables directly into the linear model. As expected, a significant negative effect is found between air pollution and self-report authenticity. We also construct the corresponding binary index with 1 being not completely true and 0 being completely true. Column (5) re-runs the probit model, and reports the marginal effect.

Overall, although the results vary a bit in magnitude across different measures, they all indicate a clear negative effect of air pollution on people's attention.

7. Placebo tests

In this section, referring to Heyes and Zhu (2019), we further conduct two placebo tests in Table 16 by replacing our matched air pollution with other irrelevant data in

time and space. Column (1) reprints the preferred estimate displayed in column (5) of Table 3. Column (2) re-runs the identical regressions under full controls based on the air pollution in reverse chronological order. Column (3) scrambles the air pollution data within our sample of cities in reverse-alphabetic order. Both of the results in column (2) and column (3) break down. It is apparent from the table that only properly correlated air pollution data in column (1) have a significant negative effect on the participants' attention ability.

8. Conclusion

Paying attention is one of the most important brain-based cognitive functions, and its disturbance has been correlated to a variety of problems in both work and life. To date, the detrimental effect of air pollution on brain health has been largely studied by way of experiments in Neuroscience. In this paper, we provide a novel way to explore the potential decrease in people's attention due to the elevated air pollution by observing the inattentive responses in online surveys.

We follow the way in Maniaci and Rogge (2014) and construct the inconsistent index to measure the inattentive behaviour of the participants in online surveys. The corresponding daily average air pollution is connected to each observation according to the disclosed IP address and the responding time. The estimated effect is substantial and robust under a set of controls. In our preferred specification, a one standard deviation of PM2.5 (around 30 $\mu g/m^3$) increases the likelihood of inconsistency by 2.25%. Taking into account the heterogeneous effects of air quality, we further examine the regressions under different demographic levels. It seems that air pollution appears to have larger effects on those young groups with high education but low income, especially for students, and the number is 5.36%.

Exploring the polluted attention existing in online surveys is an important outcome in its own right, helping to enhance the reliability and statistical power, and also provides a mechanism to support the recent study about air pollution and productivity.

Notes

1. Dr. Cory-Slechta is a well-known expert on neurotoxicology and psychology at the University of Rochester Medical School (URMC).
2. The detail description of the infrequency scales and the inconsistency scales will be explained in the following data section.
3. 50.6% of the respondents have household income of $6,123.60 to $15,309.01 if converting Chinese Yuan to US Dollar.
4. Neutrality is a tricky issue, and the most direct way to deal with that is to ignore any potential inconsistency when people reply neutral, in which the effect of air pollution is estimated to be statistically significant. However, after our in-depth thinking about the question designed for the respondents in 2015 and 2017, neutrality in this case is viewed as a positive sense of social security.

   As for the first question, the words of "chaotic", "major disturbance" are strongly negative, which is in fact obviously inconsistent with the national conditions at that moment. If the respondents keep neutral and do not fully agree with it, they might think

that there are some problems in the society, but it is less likely to fall into the extreme slumps, which just fits with the part of contents in another question. Therefore, neutrality can be seen as a positive attitude here. For the second question, the statement is positive. Neutrality to a positive question can also be considered as a positive attitude.

5. There were 1598 monitoring sites established by the end of 2017.
6. The IP address is an "identity card" obtained based on each computer's network connection point. Querying the IP address allows us to effectively locate the user's city.
7. All the data analysis and regressions are performed using Stata 16.0.
8. PM2.5 larger than 250 $\mu g/m^3$ is ignored for the lack of enough data in our sample.
9. The effect size (2.25%) can be computed by multiplying the standard deviation of PM2.5 (29.9028 $\mu g/m^3$) and the average marginal effect (0.0751%).
10. Worker includes the jobs in agriculture, forestry, animal husbandry, Fishery, and factory.

Disclosure statement

Funding

ORCID

Mingying Zhu 🆔 http://orcid.org/0000-0002-3423-074X

References

Ao, C.-K., Dong, Y., & Kuo, P.-F. (2021). Industrialization, indoor and ambient air quality, and elderly mental health. *China Economic Review*, *69*, 101676. https://doi.org/10.1016/j.chieco.2021.101676

Balakrishnan, U., & Tsaneva, M. (2021). Air pollution and academic performance: Evidence from India. *World Development*, *146*, 105553. https://doi.org/10.1016/j.worlddev.2021.105553

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, *111*(2), 218–229. https://doi.org/10.1037/pspp0000085

Brühlmann, F., Petralito, S., Aeschbach, L. F., & Opwis, K. (2020). The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology*, *2*, 100022. https://doi.org/10.1016/j.metip.2020.100022

Butcher, J. N. (2010). Minnesota multiphasic personality inventory. *The Corsini Encyclopedia of Psychology*, 1–3. https://doi.org/10.1002/9780470479216.corpsy0573

Chang, T. Y., Graff Zivin, J., Gross, T., & Neidell, M. (2019). The effect of pollution on worker productivity: Evidence from call center workers in China. *American Economic Journal: Applied Economics*, *11*(1), 151–172. https://doi.org/10.1257/app.20160436

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology*, *32*(4), 347–361. https://doi.org/10.1007/s10869-016-9458-5

Choi, Y.-R., & Chun, C.-Y. (2009). The effect of indoor temperature on occupants' attention abilities. In *Proceeding of Spring/Autumn Annual Conference of KHA* (pp. 199–203). The Korean Housing Association.

Cobanoglu, C., Moreo, P. J., & Warde, B. (2001). A comparison of mail, fax and web-based survey methods. *International Journal of Market Research*, *43*(4), 1–15. https://doi.org/10.1177/147078530104300401

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Elder, A., Gelein, R., Silva, V., Feikert, T., Opanashuk, L., Carter, J., Potter, R., Maynard, A., Ito, Y., Finkelstein, J., & Oberdörster, G. (2006). Translocation of inhaled ultrafine manganese oxide particles to the central nervous system. *Environmental Health Perspectives*, *114*(8), 1172–1178. https://doi.org/10.1289/ehp.9030

Graff Zivin, J., & Neidell, M. (2012). The impact of pollution on worker productivity. *The American Economic Review*, *102*(7), 3652–3673. https://doi.org/10.1257/aer.102.7.3652

Grineski, S. E., Clark-Reyna, S. E., & Collins, T. W. (2016). School-based exposure to hazardous air pollutants and grade point average: A multi-level study. *Environmental Research*, *147*, 164–171. https://doi.org/10.1016/j.envres.2016.02.004

He, J., Liu, H., & Salvo, A. (2019). Severe air pollution and labor productivity: Evidence from industrial towns in China. *American Economic Journal: Applied Economics*, *11*(1), 173–201. https://doi.org/10.1257/app.20170286

Heyes, A., & Zhu, M. (2019). Air pollution as a cause of sleeplessness: Social media evidence from a panel of Chinese cities. *Journal of Environmental Economics and Management*, *98*, 102247. https://doi.org/10.1016/j.jeem.2019.07.002

Jew, K., Herr, D., Wong, C., Kennell, A., Morris-Schaffer, K., Oberdörster, G., O'Banion, M. K., Cory-Slechta, D. A., & Elder, A. (2019). Selective memory and behavioral alterations after ambient ultrafine particulate matter exposure in aged 3xTgAD Alzheimer's disease mice. *Particle and Fibre Toxicology*, *16*(1), 1–17. https://doi.org/10.1186/s12989-019-0323-3

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Kilian, J., & Kitazawa, M. (2018). The emerging risk of exposure to air pollution on cognitive decline and Alzheimer's disease-Evidence from epidemiological and animal studies. *Biomedical Journal*, *41*(3), 141–162. https://doi.org/10.1016/j.bj.2018.06.001

Lee, H., Choi, Y., & Chun, C. (2012). The effect of indoor air temperature on occupants' attention ability based on the electroencephalogram analysis. In *10th International Conference on Healthy Buildings 2012* (pp. 2083–2088).

Liu, M., Tan, R., & Zhang, B. (2021). The costs of "blue sky": Environmental regulation, technology upgrading, and labor demand in China. *Journal of Development Economics*, *150*, 102610. https://doi.org/10.1016/j.jdeveco.2020.102610

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Morey, L. C., & Lowmaster, S. E. (2010). Personality assessment inventory. *The Corsini Encyclopedia of Psychology*, 1–4. https://doi.org/10.1002/9780470479216.corpsy0663

Peeples, L. (2020). News Feature: How air pollution threatens brain health. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(25), 13856–13860. https://doi.org/10.1073/pnas.2008940117

Pilcher, J. J., Nadler, E., & Busch, C. (2002). Effects of hot and cold temperature exposure on performance: A meta-analytic review. *Ergonomics*, *45*(10), 682–698. https://doi.org/10.1080/00140130210158419

Qin, Y., & Zhu, H. (2018). Run away? Air pollution and emigration interests in China. *Journal of Population Economics*, *31*(1), 235–266. https://doi.org/10.1007/s00148-017-0653-0

Rice, S., Winter, S. R., Doherty, S., & Milner, M. (2017). Advantages and disadvantages of using internet-based survey methods in aviation-related research. *Journal of Aviation Technology and Engineering*, *7*(1), 5. https://doi.org/10.7771/2159-6670.1160

Sager, L. (2019). Estimating the effect of air pollution on road safety using atmospheric temperature inversions. *Journal of Environmental Economics and Management*, *98*, 102250. https://doi.org/10.1016/j.jeem.2019.102250

Stoerk, T. (2016). Statistical corruption in Beijing's air quality data has likely ended in 2012. *Atmospheric Environment*, *127*, 365–371. https://doi.org/10.1016/j.atmosenv.2015.12.055

Tuten, T. L., Urban, D. J., & Bosnjak, M. (2002). Internet surveys and data quality: A review. *Online Social Sciences*, *1*, 7–26.

Wan, Y., Li, Y., Liu, C., & Li, Z. (2020). Is traffic accident related to air pollution? A case report from an island of Taihu Lake, China. *Atmospheric Pollution Research*, *11*(5), 1028–1033. https://doi.org/10.1016/j.apr.2020.02.018

Yeganeh, A. J., Reichard, G., McCoy, A. P., Bulbul, T., & Jazizadeh, F. (2018). Correlation of ambient air temperature and cognitive performance: A systematic review and meta-analysis. *Building and Environment*, *143*, 701–716. https://doi.org/10.1016/j.buildenv.2018.07.002

Younan, D., Petkus, A. J., Widaman, K. F., Wang, X., Casanova, R., Espeland, M. A., Gatz, M., Henderson, V. W., Manson, J. E., Rapp, S. R., Sachs, B. C., Serre, M. L., Gaussoin, S. A., Barnard, R., Saldana, S., Vizuete, W., Beavers, D. P., Salinas, J. A., Chui, H. C., … Chen, J.-C. (2020). Particulate matter and episodic memory decline mediated by early neuroanatomic biomarkers of Alzheimer's disease. *Brain*, *143*(1), 289–302. https://doi.org/10.1093/brain/awz348

Zhang, J. (2019). Cognitive functions of the brain: Perception, attention and memory. ArXiv, abs/1907.02863. https://doi.org/10.48550/arXiv.1907.02863