scientific reports

OPEN

Health risk assessment of PM$_{2.5}$ on walking trips

Caihua Zhu, Zekun Fu, Linjian Liu, Xuan Shi & Yan Li[✉]

PM$_{2.5}$ has an impact on residents' physical health during travelling, especially walking completely exposed to the environment. In order to obtain the specific impact of PM$_{2.5}$ on walking, 368 healthy volunteers were selected and they were grouped according to gender and age. In the experiment, the heart rate change rate (HR%) is taken as test variable. According to receiver operating characteristic (ROC) curve, the travel is divided into two states: safety and risk. Based on this, a binary logit model considering Body Mass Index (BMI) is established to determine the contribution of PM$_{2.5}$ concentration and body characteristics to travel risk. The experiment was conducted on Chang'an Middle Road in Xi'an City. The analysis results show that the threshold of HR% for safety and risk ranges from 31.1 to 40.1%, and that of PM$_{2.5}$ concentration ranges from 81 to 168 μg/m^3. The probability of risk rises 5.8% and 11.4%, respectively, for every unit increase in PM$_{2.5}$ concentration and HR%. Under same conditions, the probability of risk for male is 76.8% of that for female. The probability of risk for youth is 67.5% of that for middle-aged people, and the probability of risk for people with BMI in healthy range is 72.1% of that for non-healthy range. The research evaluates risk characteristics of walking in particular polluted weather, which can improve residents' health level and provide suggestions for travel decision while walking.

The health risks in residents' travel are mainly related to the concentration of particulate matter (PM)[1]. PM is the primary cause of many adverse health effects, including but not limited to cardiovascular and respiratory diseases[2,3]. As the main PM exposure environment for residents, traffic environment involves a large number of people. PM has a more significant impact on commuters with intensive travel time and high travel demand[4]. Walking as the final transfer mode in a trip is an indispensable part to complete the travel. Since walking is fully exposed to the external environment, it is very sensitive to the change of PM concentration. In order to assess the health risk of walking trips, we need to address two aspects: (1) The human body has its own regulatory capacity. How to determine the boundary between the safe state and the risk state? (2) The quantitative description of the factors that influence the health risk of travel is not clear. If the health risk degree at certain PM concentration could be obtained, the health risk degree, i.e., the probability of health risks when residents travel in specific polluted weather. For commuters, it will be possible to alter residents' willingness to travel or their travel modes, and improve the health conditions in travelling. For managers, it will provide risk assessment basis for health control, and make commuters' travel reach the overall optimal level.

The health risks during commuting are not only closely relevant to the concentration of pollutants in the external environment, but also connected to residents' personal physical conditions. Due to people's different physical qualities, their resistance to pollution is not always the same[5]. Individual characteristics (age, gender, figure and so on) have an effect on physical quality[6,7]. When estimating the critical value of pollutant concentration corresponding to safety and risk, individual situations of each volunteer should be analyzed. The health risk of walking trips in polluted weather can be reflected by a variety of factors. Among them, real-time physiological data during travel is the most active reflection to the disturbance of external environment. Heart rate as the most intuitive sample data has high reliability since it is less influenced by commuters' subjective consciousness[8]. Thus, the impact of pollutant concentration on residents' health can be measured by the changes of heart rate. The health status is affected by a combination of the external environment and physical condition. So, it is necessary to study the specific contribution of each influencing factor to health risk. Because the experiments were conducted simultaneously, the analysis was carried out from the perspective of PM concentration. The novelty of this study lies in obtaining health risk thresholds for volunteers in terms of heart rate and quantitatively describing the relationship between each influencing factor and health risk degree.

The sources of PM include natural and anthropogenic sources. Natural sources mainly include dust storms, volcanic eruptions, forest and grassland fires; anthropogenic sources mainly include traffic trips, home heating, power plants and various industrial processes[5]. Due to the numerous natural and anthropogenic sources, PM may

College of Transportation Engineering, Chang'an University, Middle section of south 2nd Ring Road, Xi'an 710064, Shaanxi Province, China. ✉email: lyan@chd.edu.cn

Study area (place)	Pollutant types	Key observations	Author (year)
Chile	$PM_{2.5}$	Personal PM exposure concentration and its influencing factors of commuters with different transportation modes	Suárez, L. et al. (2014)
Iran	$PM_{2.5}/PM_{10}$	Concentrations of annual PM exceeding the WHO air quality guideline, and an unacceptably high risk for human health	Yunesian, M. et al. (2019)
China	$PM_{2.5}$	Short-term exposure to ambient PM2.5 was significantly associated with an increased risk of daily outpatient visits for ulcerative colitis, and related to gender and age	Duan, R. et al. (2021)
Iran	PM_{10}	The average PM10 concentration was higher in summer. Higher exposure levels in female	Ahmadi, S. et al. (2021)
Iran	PM	Most of particles were inorganic in nature, and PM may have different physicochemical characteristics in different areas	Sajjadi, S. A. et al. (2018)
Iran	$PM_{2.5}/PM_{10}$	The PM concentration was higher in the warm season than in the cool season, and the number of colonies increased with the increase in PM concentration	Amarloei, A. et al. (2020)
India	$PM_1/PM_{2.5}/PM_{10}$	PM concentrations are accompanied by spatial shifts that are related to the frequency of human activity	Sahu, V. et al. (2018)
India	$PM_{2.5}/PM_{10}$	The number of hospitalizations for respiratory problems shows a positive correlation with PM concentrations, and PM10 has 2 times more impact on human health than PM2.5	Gupta, A. et al. (2019)
China	$PM_{2.5}$	The spatial distribution of PM2.5 concentration in Xi'an and the building distribution does not match	Sun, X. et al. (2020)

Table 1. Summary of relevant past the characteristics of PM concentration distribution and the relationship between PM concentration and health risk.

have different physicochemical characteristics in different areas and it is diffuse in nature[5,9]. PM concentrations show seasonality, with higher average PM concentrations in the warm season than in the cool season, and higher PM concentrations from June to September[7,10]. Meanwhile, factors such as distance from dust source, frequency and duration of dust storms, vegetation cover, and soil type also have a greater influence on PM concentrations[7]. In areas of human activity, PM concentrations are accompanied by spatial shifts that are related to the frequency of human activity[11]. The increased concentration of PM to which individuals are exposed may alter metabolic activity and lead to serious diseases such as asthma, rhinitis, and tuberculosis[12]. The number of hospitalizations for respiratory problems shows a positive correlation with PM concentrations[13].

PM concentrations are generally collected with the help of sampling instruments to monitor individual exposure levels[14]. There are three main methods in the research of PM concentration. The first method is to estimate the impact by measuring PM concentration in the environment. It is a kind of intuitive statistics with single environmental variable data, which is often used to compare PM concentrations in different regions[15]. The second method is to establish a dose model through breathing frequency to estimate the suction volume of pollutant. Then the size of suction volume is adopted to determine the degree of risk. This type of method is often used in the comparison of different travel modes[16,17]. The last method is to evaluate the impact on human health by combining PM concentrations with physiological indicators. Most of collected statistics are time series data, which can illustrate the variation tendency of physiological data with PM concentration changes[18,19]. Commonly used physiological indicators are electrocardiogram (ECG) and electroencephalogram (EEG)[20]. ECG is more widely used due to its simple and accurate measurement[21]. Heart rate and heartbeat interval[22] are highly stable and they could visualize changes in the external environment. So, these two indexes are usually applied to describe health effects caused by PM concentrations.

In the statistics of travel health risk, PM concentrations are main considerations in the degree of risk. The higher the PM concentrations, the greater the risk. Common approaches in identifying risks include intuitive statistics[23], Markov model[24], Strengths Weaknesses Opportunities Threats (SWOT)[25], Receiver Operating Characteristic (ROC) curves[26] and so on. Judging health risk actually lies in the determination of threshold for safety and risk in the travel. ROC curves as a series of different binary classification methods are able to calculate critical value of risk. High classification accuracy of ROC curves leads to their widespread application in the field of medical image recognition and disease recognition[27]. Intuitive analysis models for PM focus on linear probability models and discrete probability models. The former is mainly used to connect PM to health indicators[28]. The latter is adopted to calculate the probability of a specific event, among which Logit model and Probit model are representative models. Logit model has a faster solution speed as a derivative of discrete choice models[4]. This research is based on the analysis of binary classification for safety and risk, so binary logit model can be selected to model and analyze influencing factors of health during travelling. Previous studies (Table 1) have shown that PM has an impact on human health, and this impact is related to exposure to the environment and physical conditions.

PM concentrations have an impact on the health of residents who travel walking. However, PM concentration thresholds that create risk and the quantitative description of associated influencing factors remain unknown. The objective of this study is to obtain the probability of health risk caused by walking in polluted weather based on actual experimental investigations. Major contributions of the research include: (1) the health risk threshold of each volunteer is determined with heart rate indicators as input variables; (2) binary logit model of health risk is established with safety and risk as defined interval, and the relationship between influencing factors and health risk is quantitated, the factors that contributed most to health risk are determined. The paper adds to the current literature by quantitatively describing the effects of PM on human health. This study adds a health risk threshold jointly determined by PM concentration and heart rate indicator, and a quantitative description of the relationship between influencing factors and health risk with the help of a dichotomous approach, providing basic theoretical knowledge for the detection of health risks in other travel modes.

Methods

Framework.
Commuters' physical functions vary from person to person, and the same environmental factors have different effects on human health. So, it is necessary to make a concrete analysis considering commuters' own situations. PM concentration will not impact health extensively in a certain range due to the self-adjustment function of human body. Only if it is higher than a fixed value can this impact be reflected. Therefore, it is required to find out this critical value in the analysis of health risk.

Identifying safe and risk state belongs to binary classification problem. ROC curve as a binary classification machine learning algorithm can intuitively reflect measurement accuracy of the model at different thresholds and effectively solve binary classification problem. Note that the process of determining thresholds between safe and risk state is conducted for each volunteer independently. While studying the effect of PM concentration on travel health, relying solely on qualitative descriptions has no actual meaning. This raises the need for a relationship model between PM concentration and health risk. According to health risk thresholds, safe state and risk state could be defined during travelling. Then binary logit model can be applied in establishing the relationship model to quantitatively obtain residents' travel health risk under different PM concentrations.

ROC curve.
ROC curve is applicable in binary classification cases, that is, to judge two kinds of results "yes" or "no". ROC curve is a kind of curve that describes the relationship between sensitivity (or True Positive Rate, TPR) and specificity (or False Positive Rate, FPR) with a conformational method in the test[29]. AUC (Area Under ROC Curve) reflects the accuracy of threshold and its value ranges from 0.5 to 1.0. When AUC is greater than 0.9, the accuracy is high. Each point on ROC curve is obtained by changing the classification thresholds in the same model. Thus, the process of determining the best point on the same ROC curve is exactly the process of determining the best classification threshold. If the point is near to the upper left, it has better discriminative effect[30]. In the application, the point with the largest Youden index (Youden index = sensitivity + specificity—1) can be selected as the discriminative threshold on the curve.

In this study, the experimental process is organized as follows:

(1) According to the subjective judgment by volunteers after the experiment, PM concentrations are divided into initial safe concentration and risk concentration as the state variable of the model.
(2) With HR% as the test variable, basic data points are selected. Then the threshold interval in each calculation is estimated according to the input HR%, and the initial value vector is constructed.
(3) Based on discrimination results, Z_{TP} (risk samples and judged as risk samples; number of correctly affirmed matches), Z_{FN} (risk samples but judged as safe samples; number of misses, not correctly found matches), Z_{FP} (safe samples but judged as risk samples; false positives, number of incorrect matches), Z_{TN} (safe samples and judged as safe samples; number of correctly rejected non-matches) corresponding to each threshold are calculated.
(4) The values of sensitivity (TPR) and specificity (FPR) are calculated, where $T = Z_{TP}/(Z_{TP} + Z_{FN})$,$F = Z_{FP}/(Z_{FP} + Z_{TN})$.
(5) Connect the points (F, T) of each threshold. Then plot ROC curve and calculate Youden index. The feature point with the largest Youden index on the upper left can be distinguished. HR% corresponding to this feature point is the discriminative threshold of health risk.

Model construction considering BMI.
As a probabilistic model, logit model can predict the probability of an event[31]. This research is related to a binary-state variable reflecting safe and risk state during travelling, so it can be analyzed by binary logit model. In binary logit model, the dependent variable can only take two values 1 and 0 (virtual dependent variable). In this study, the model can not only estimate whether a volunteer is in a risk state, but also calculate the degree of risk at that time. The model are formulated as:

$$\hat{Y} = \ln \frac{p}{1-p} = \beta_0 + \sum_{i=1}^{n} \beta_i X_i, i \in N \tag{1}$$

where, $\hat{Y}$ is the virtual dependent variable and its value is 0 or 1; p is the probability of risk and its value ranges from 0 to 1; β_0 is the undetermined coefficient; β_i is the coefficient corresponding to the i-th explanatory variable; X_i is the i-th explanatory variable.

The operation of model consists of two parts. The first part is to solve the coefficient of explanatory variable according to virtual dependent variable $\hat{Y}$. The second part is to deduce the probability of risk p based on the variable coefficient and the value of explanatory variable. The probability of risk is calculated as:

$$p = \exp\left(\beta_0 + \sum_{i=1}^{n} \beta_i X_i\right) \bigg/ \left[1 + \exp\left(\beta_0 + \sum_{i=1}^{n} \beta_i X_i\right)\right] \tag{2}$$

Commuters will eventually turn to different risk states due to their different physical conditions even in the same environment. Body Mass Index (BMI) is a commonly used standard to measure the degree of obesity and health. For comparing and analyzing the health effects of weight on people with different heights, BMI is a neutral and reliable index.

In the process of experiment, BMI index is introduced as one of explanatory variables to measure volunteers' health information. This approach aims to avoid the impact of health disparities on model accuracy. The calculation formula of BMI is shown as:
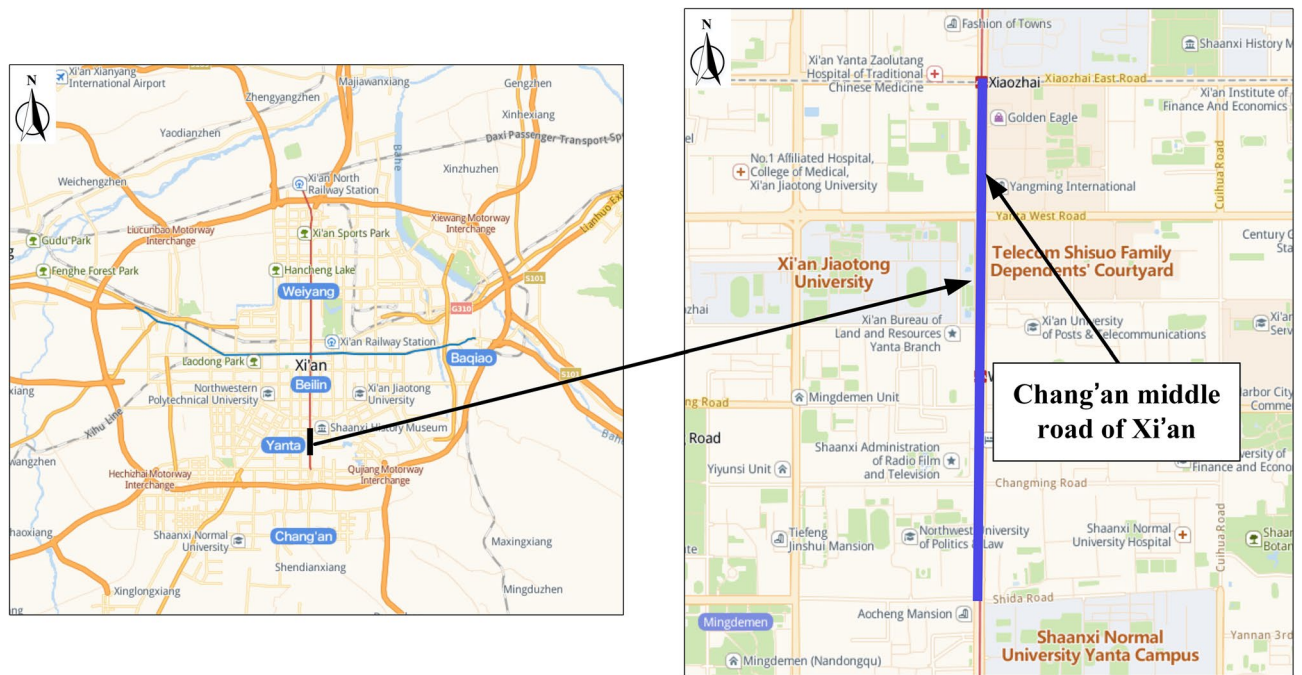
Figure 1. Location of experimental path and its details. And figure was plotted by Microsoft Office Visio 2013, which can be downloaded on https://www.microsoft.com.

$$BMI = \frac{BW}{H^2} \tag{3}$$

where: BW is the weight with unit of kg; H is the height with unit of m. For healthy people, their BMI should be in the range of 18.5 to 24.

Experiments

Experimental design and path selection.   In all procedures, we adhered to the declaration of Helsinki, and the study was received ethics approval from the Ethics Committee of Chang'an University. All participants are over 18 years old. We obtained informed consent from all study participants who were enrolled in the study, and participants were free to leave the experiment at any stage if they felt uncomfortable.

All experiments in this research were conducted during peak hours (7:30 to 8:30 and 17:30 to 18:30) from September to December (fall and winter) in 2019. The research objects were travel processes of volunteers. Since weather had a great influence on the assessment of air pollution concentrations, no tests were arranged in rainy and snowy days. In the selection of volunteers, they were required to be physically and mentally healthy without history of cardiovascular disease, pulmonary disease, nervous system disease and some others. Taking medicines was also not allowed. Then all candidates had a full medical examination to prevent data error caused by physical diseases. Finally, 368 healthy volunteers were selected for the test. All volunteers were grouped according to age and gender. On the basis of classification criteria proposed by World Health Organization[32], volunteers were divided into 112 young men, 97 young women, 85 middle-aged men and 74 middle-aged women. All participants should rest well in the night before the experiment. Tobacco, wine, tea, coffee, any other food or drugs that might affect volunteer physical status were forbidden in the previous day. Walking is assigned as their travel mode. Before the experiment, the heart rate under relaxed state was measured for 10 min in an air-purified space, which is considered as a healthy heart rate in a pollution-free environment. At the end of experiment, volunteers were asked to complete a subjective questionnaire to determine whether they stayed in risk state at a certain PM concentration.

Middle Chang'an Road in Xi'an, China, was chosen as a typical experimental path where motor and non-motor vehicles are separated and exclusive sidewalks are set (Fig. 1). The experiment was conducted on the sidewalk. Located in the southern part of Xi'an, this arterial road passes through Xiaozhai business district, high-density residential areas and educational areas. The total length of experimental path is 2 km. The widths of sidewalks range from 3 to 6 m. During the test, average motor vehicle volume on this section was 2662 vehicles/h, consisting of 79% private cars, 12% taxis, 7% buses and less than 2% trucks. The volume of non-motor vehicles was 3008 vehicles/h, containing 42% electric vehicles and 58% bicycles. Pedestrian flow was 1127 people/h.

Equipment.   CEM-DT96 portable particle detector, which supports dual mode of collecting particle concentration and mass concentration, is adopted to detect PM concentration. The sampling interval is set to 1 min considering that walking speed is relatively slow. Each volunteer wears a GARMIN heart rate band with an external GPS module. The equipment continuously records heart rate data and time data in seconds. GARMIN heart
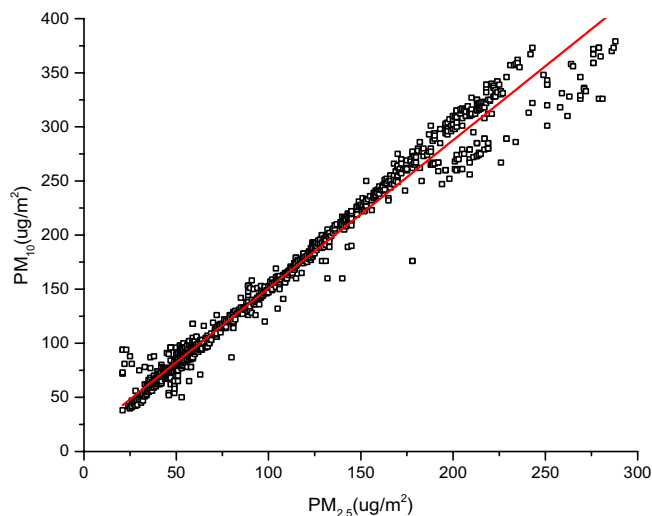
Figure 2. Scatter diagram of relationship between $PM_{2.5}$ and PM_{10}.

Time interval	30 s	60 s	120 s	180 s	240 s	300 s
Sig. F	0.009	0.011	0.027	0.062	0.115	0.136
F	126.3	62.1	58.3	52.0	47.5	38.2

Table 2. Sig. F and F values at different time intervals.

rate band is composed of thin film electrode and fabric band, which is worn under the chest line and stick to the skin. Compared with electronic bracelet, GARMIN heart rate band has higher accuracy. In order to guarantee the accuracy and effectiveness of data, the equipment needs to be calibrated synchronously in the same environment before each experiment. All time points take GARMIN detector as reference to ensure the consistency of data and facilitate subsequent data processing.

In the experiment, all volunteers wore GARMIN heart rate band tuned to uniformity and held a qualified CEM-DT96 portable particle detector in their hands to measure heart rate and PM concentration during the walking condition. Heart rate data can be directly exported from the heart rate band, and PM concentration is obtained by directly recording data on the portable particle detector.

Data processing. Collected PM concentrations, including $PM_{2.5}$ concentration and PM_{10} concentration, were derived from the particulate matter detector. Note that sampling time is supposed to correspond to that of GARMIN heart rate band. $PM_{2.5}$ and PM_{10} data represents pollutant concentrations in the external environment. $PM_{2.5}$ and PM_{10} retain a highly positive correlation, as illustrated in Fig. 2. The particle diameter of $PM_{2.5}$ is smaller, so it is easy to be affected by air buoyancy and difficult to sink[33]. As a result, its impact on the atmospheric transparency lasts longer. At the same time, $PM_{2.5}$ can enter the lungs and lead to the reaction of body function, which is more harmful to human than PM_{10}[34]. Therefore, $PM_{2.5}$ is analyzed in the modeling of health risk.

The changes of heart rate reflect subtle changes of people in different conditions, which is the basic performance of external environment acting on human. Consequently, it is essential to analyze the relationship between the changes of heart rate and $PM_{2.5}$ concentration. In data statistics, indicators related to heart rate are calculated in seconds.

Results
Data statistics.
In statistical analysis of data, data set interval has an impact on the accuracy of results and determines the workloads of data processing. If the time interval is too small, the heart rate data will have greater randomness, which tends to highlight some random fluctuations. On the contrary, large intervals may bring about inadequate display of the overall trend and lower accuracy of heart rate changes. As shown in Table 2, significance F (sig. F) and F value reflected in the changes of heart rate data among different samples can be used to find out the appropriate time interval. In statistics, sig. F is the probability that the null hypothesis cannot be rejected in proposed model. F value is a ratio computed via dividing average regression sum of squares by average error sum of squares. The smaller the sig. F (should be less than 0.05) and F value are, the better performance will be. Through integrated into account, the time interval is determined to be 60 s in data processing.

For each group of volunteers, the relationship between heart rate and $PM_{2.5}$ concentration is exhibited in Fig. 3. With the increase of $PM_{2.5}$ concentration, the heart rate has an overall upward trend despite some fluctuations. For male youth group, the heart rate is significantly lower than other groups and increases linearly. The
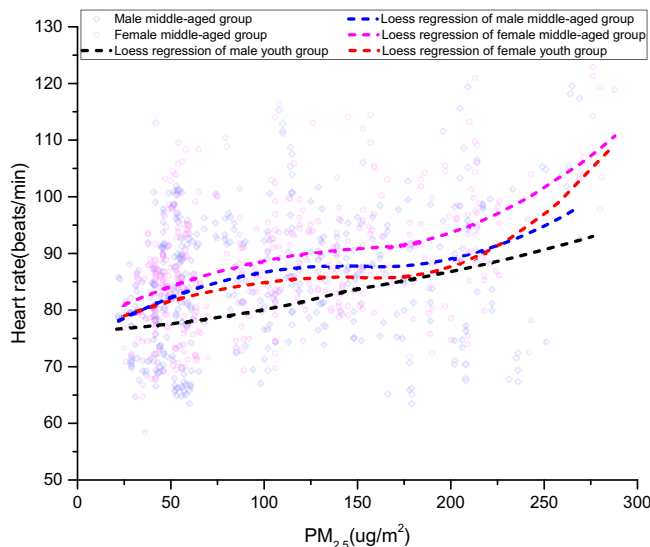
Figure 3. Variation tendency of heart rate.

Group	Mean	Standard deviation	Maximum	Minimum
Male youth group	80.69	10.86	112	58
Female youth group	84.14	10.29	125	59
Male middle-aged group	85.29	9.78	124	60
Female middle-aged group	88.25	9.18	126	59

Table 3. Statistics of heart rate indexes.

heart rate in female middle-aged group is the highest. Its growth trend is similar to that of female youth group and male middle-aged group with an "S" curve. Two inflection points on the curve represents $PM_{2.5}$ concentrations of 100ug/m^2 and 180 ug/m^2, respectively. The regression curves for male middle-aged group and female youth group have an intersection point. The heart rate of male middle-aged group is higher than that of female youth group when $PM_{2.5}$ concentration is below 220 ug/m^2. While $PM_{2.5}$ concentration is above 220 ug/m^2, the reverse applies. On the whole, youths are better at regulating the pollutant effects on body function. The physical performance of males are more stable than females as $PM_{2.5}$ concentrations increase.

Table 3 shows the heart rate statistics of volunteers in each group. The average heart rate in female middle-aged group is about 5% higher than that of female youth group and about 9% higher than male youth group. The average heart rate is approximately equally represented in female youth group and male middle-aged group with 1% gap. Middle-aged volunteers maintain a higher heart rate and smaller room to adjust in the polluted environment. In contrast, youths appear better adaptability which is represented by higher standard deviations, i.e., larger adjustment room.

Determination of health risk thresholds. In the experiment, the heart rate varied from person to person due to differences in basic physical conditions and their status on that day. To avoid this kind of influence, HR% is adopted to characterize heart rate changes for different volunteers and different experiment dates. 10 min before the experiment, the heart rate was measured in a room with air purification and its mean value was taken as the heart rate in healthy state. HR% is expressed by the formula (4):

$$r_i = \frac{t_i - t}{t} \tag{4}$$

where: r_i is the average change of heart rate during time interval i; t is the average heart rate measured 10 min before the experiment; t_i is the average heart rate during time interval i.

In this research, ROC curve is applied in order to calculate the threshold of health risk. The input state variable is HR% and the test variable is subjective judgment of safety and risk for volunteers. Safe state takes 0 as the value and 1 represents risk state. Possible thresholds should be selected at first, then the sensitivity and specificity of each threshold are calculated. Based on the value range of HR%, 0.005 is determined as the discriminant interval of threshold. For 337 out of 368 volunteers, the area under ROC curve is larger than 0.9, indicating high accuracy of the experiment. With the increase of $PM_{2.5}$ concentration, HR% gradually rises. The existing of fluctuations leads to the situation that HR% at the optimal critical point may correspond to multiple $PM_{2.5}$ concentration

(**a**) risk threshold for PM$_{2.5}$ concentration (**b**) risk threshold for HR%
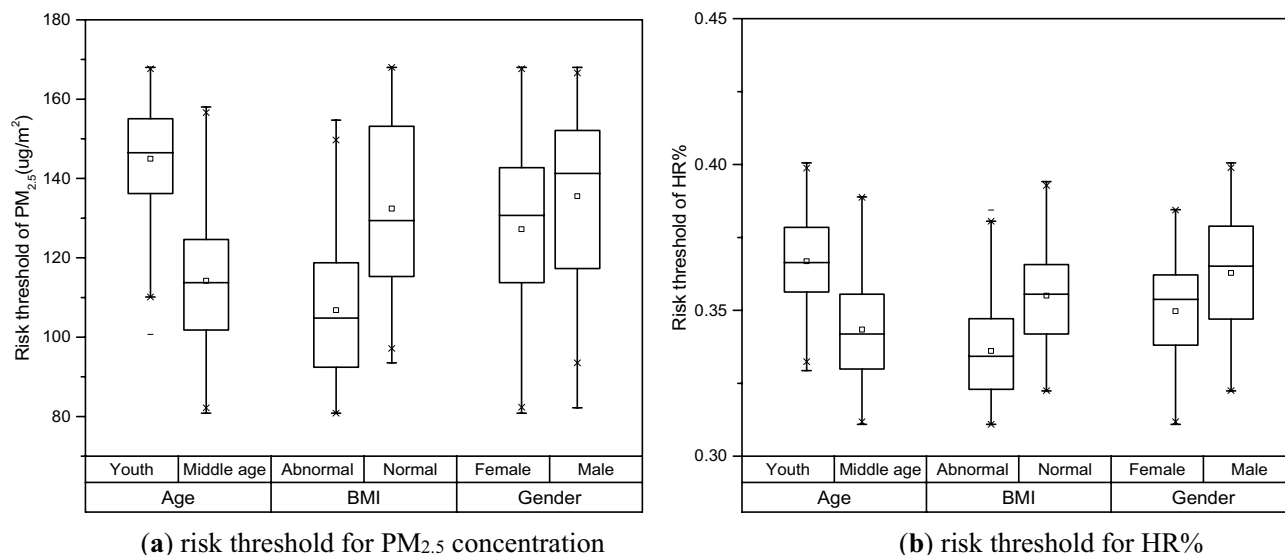
Figure 4. Distribution of health risk thresholds among volunteers.

	HR%	PM$_{2.5}$	Gender	Age	BMI
HR%	1	0.386	0.337	0.261	0.355
PM$_{2.5}$	0.386	1	0.053	0.021	0.102
Gender	0.337	0.053	1	0.034	0.064
Age	0.261	0.021	0.034	1	0.217
BMI	0.355	0.102	0.064	0.217	1

Table 4. Correlation of explanatory variables. **The correlation was significant at the 0.01 level (two-tailed).

values. The solution is to take the average of these values as the threshold for safe and risk state. The thresholds for volunteers under different physical conditions are shown in Fig. 4.

HR% at the optimal threshold of health risk ranges from 31.1% to 40.1%, corresponding to a PM concentration range of 81 to 168 ug/m^2. For male volunteers, the average concentration at which risk occurs is 136 ug/m^2. It is higher than that for females (127 ug/m^2), indicating that males show greater resistance to air pollution. The risky concentration for youth is 145 ug/m^2, significantly higher than 114 ug/m^2 for middle aged people. As people age, the hypofunction results in lower receptivity to the external environment. If a volunteer's BMI stays in the normal (healthy) section, his (her) risky PM concentration is 132 ug/m^2, which is higher than 107 ug/m^2 in abnormal (non-healthy) section. It proves that physical condition will affect the reaction to the external environment. Good physical quality would increase the adaptability and improve pollution resistance.

Health risk characteristics. The occurrence of health risks is correlated with HR%, pollution exposure concentration, gender, age and BMI. In order to avoid the error caused by high correlation of these explanatory variables, a correlation test should be undertaken before modeling. Pearson correlation coefficient was utilized in this research. Pearson correlation coefficient takes values between 0 and 1. It can be divided into five intervals with an interval step of 0.2. The correlation coefficient (from large to small) indicates the correlations are very strong, strong, moderate, weak and very weak. Correlations between different explanatory variables are listed in Table 4.

As illustrated in Table 4, there is no significant correlation between explanatory variables. Thus, five variables (PM$_{2.5}$ concentration, gender, age, BMI and HR%) can be selected as explanatory variables. The occurrence of health risks is correlated with HR%, pollution exposure concentration, gender, age and BMI. According to the threshold of health risk, the status of volunteers is classified into two categories: safe state and risk state. Then binary logit model is established.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\gamma_1 + \beta_2\gamma_2 + \beta_3\gamma_3 + \beta_4\gamma_4 + \beta_5\gamma_5 \tag{5}$$

where: p is the health risk level; γ_1 is the concentration of PM$_{2.5}$; γ_2, γ_3, and γ_4 are 0–1 dummy variables (γ_2: 1-male, 0-other; γ_3: 1-youth, 0-other; γ_4:1-BMI in healthy section, 0-other); γ_5 is HR% with an interval of 60 s; β_0 is the undetermined coefficient; $\beta_1 \sim \beta_5$ are the coefficients of $\gamma_1 \sim \gamma_5$ variables, respectively.

Variables	β_j	Standard Error	Wald test	Degree Freedom	P value	Exp (β_j)
γ_1	0.056	0.578	68.578	1	0.000	1.058
γ_2	−0.264	0.275	6.235	1	0.023	0.768
γ_3	−0.393	0.069	29.189	1	0.008	0.675
γ_4	−0.327	0.322	12.624	1	0.011	0.721
γ_5	0.108	0.396	17.591	1	0.016	1.114
β_0	−6.323	0.821	39.467	1	0.005	0.002

Table 5. The value of model parameters (HR%, $PM_{2.5}$ concentration, gender, age and BMI). The statistical value is the mean value of the parameters.

The comprehensive analysis results for each parameter are shown in Table 5. The significance levels of all variables are less than 0.05, which proved the validity of the model. Among all coefficients of variables, β_1 and β_5 are positive and the remaining are negative. It indicates that the higher the $PM_{2.5}$ concentration and HR%, the higher the probability of generating health risks. Males, youth and those with BMI in normal section are more adaptable to polluted weather. The analysis results from model are consistent with statistical results from experiment.

The coefficient of $PM_{2.5}$ concentration is 0.056 and Exp (β_1) is 1.058, indicating that for each unit increase in concentration, the probability of risk will be 1.058 times higher. The coefficient for males is −0.264 and Exp (β_2) is 0.768, proving that the probability of risk for males is 76.8% of that for females. The coefficient for youth is −0.393 and Exp (β_3) is 0.675, indicating that the probability of risk for youth is 67.5% of that for middle aged people. The coefficient for people whose BMI falls in normal interval is −0.327 and Exp (β_4) is 0.721. It shows that the probability of risk for healthy volunteers is 72.1% of that for non-healthy ones. The coefficient of HR% is 0.108 and Exp (β_5) is 1.114. It means that for each unit increase in HR%, the probability of risk will be 1.114 times greater.

Wald test is a method to estimate the significance of a given explanatory variable in a statistical model[35]. Its value takes the ratio between the regression coefficient and the standard error. A higher Wald value means a more essential explanatory variable. In this research, $PM_{2.5}$ concentration and age have the largest Wald test values and both significance levels are less than 0.05. The explanation is that these two variables have a significant effect on the health risk degree. The Wald test value of $PM_{2.5}$ concentration is significantly higher than other variables, demonstrating that it has the greatest impact.

According to the analysis results, the health risk model is calibrated as:

$$\ln\left(\frac{p}{1-p}\right) = -6.323 + 0.056\gamma_1 - 0.264\gamma_2 - 0.393\gamma_3 - 0.327\gamma_4 + 0.108\gamma_5 \tag{6}$$

Based on formula (6), the health risk degree p can be calculated:

$$p = \frac{\exp(-6.323 + 0.056\gamma_1 - 0.264\gamma_2 - 0.393\gamma_3 - 0.327\gamma_4 + 0.108\gamma_5)}{1 + \exp(-6.323 + 0.056\gamma_1 - 0.264\gamma_2 - 0.393\gamma_3 - 0.327\gamma_4 + 0.108\gamma_4)} \tag{7}$$

The health risk degree is calculated respectively for each group of volunteers. Final results and health risk curves are shown in Fig. 5. As $PM_{2.5}$ concentrations increase, the probability of health risk gradually comes up. Striking differences are exhibited between males and females, youth and middle-aged people. It is apparent that males and youth have better resistance to pollution. At the same $PM_{2.5}$ concentration, the probability of health risk from high to low in order is: female middle-aged group > male middle-aged group > female youth group > male youth group.

Discussion

Epidemiological and toxicologic studies have shown that $PM_{2.5}$ can induce cardiovascular and respiratory diseases. Moreover, it has many other adverse effects on human health, such as Type 2 Diabetes, birth defects and premature deaths. The underlying mechanisms of the relationship between $PM_{2.5}$ and poor health include the oxidation reactions in induced cells, mutagenicity and inflammation[36]. Research results demonstrate that exposed environment with high $PM_{2.5}$ concentration will lead to increasing probability of commuters' health risks, which is consistent with former conclusions[37–39]. When the sudden change in heart rate reaches 35% or more, it means that there is a large external environmental disturbance and the body's regulatory function starts to feedback[21]. When the concentration of pollutants in the air is higher than 150 μg/m³, the body's blood oxygen saturation is below the normal range and health risks occur[40]. The above findings lie within the health risk threshold of this paper.

The youth and the commuters with BMI index in normal section have better resistance to pollution. The probability of risk for youth is 67.5% of that for middle-aged people, and the probability of risk for people with BMI in healthy range is 72.1% of that for people with BMI in non-healthy range. Compared with middle-aged people, the youth have greater capacity to adapt to the external environment. This phenomenon is mainly related to the sensitivity of body functions to external reactions. Existing researches have shown that pollutants can activate the youth's cardiorespiratory functions[41,42], and youths have 1.3 times more space for cardiorespiratory conditioning than middle-aged[43]. Both obesity and thinness have an impact on body resistance to the external
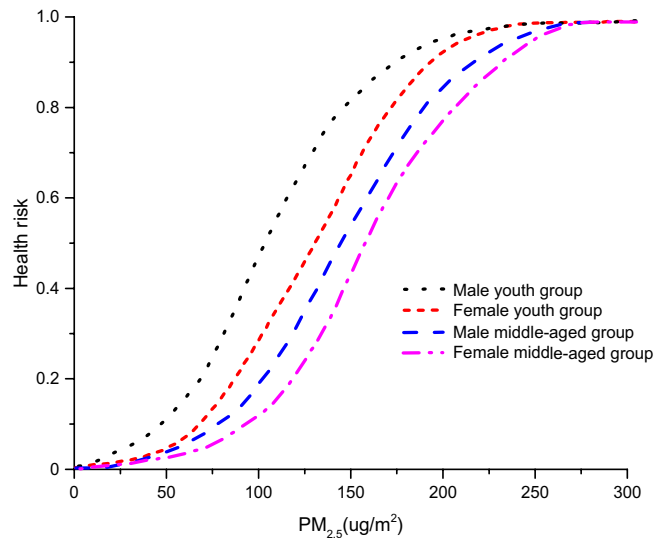
Figure 5. The probability of health risks in each group of volunteers as $PM_{2.5}$ concentration increases.

environment. In epidemiological studies, BMI index is highly correlated with anti-pollution ability. For obese people, long-term exposure to pollutants will increase their insulin production and then rise blood sugar levels[44]. People with lower BMI index tend to have poor immunity. So, their risks of coronary heart disease and hypertension caused by air pollution is 1.2 times higher than that for people with normal BMI index. As a result, the death rate is about 1.1 times higher than normal value[45,46].

An unexpected but interesting finding is the sex-specific difference of health risks under air pollution environment, which is reflected in better resistance to pollution for males. In the same travel environment, the probability of risk for male is 76.8% of that for female. This conclusion indicates that some unknown factors associated with sex or gender affect the body's response to pollutants. People are becoming more and more interested in the role of sex and gender played in the epidemiology relevant to air pollution. Several recent studies have reported that pollutants have a greater impact on female's respiratory health than male[47,48]. A comprehensive analysis of the relationship between pollutants and death rate found that in a highly polluted environment, the risk for females increased by 1.12% and that for males increased by 0.73%[49]. However, in epidemiological research, it is complicated to draw a distinction between sex (e.g., physiology, toxicokinetics) and gender (e.g., the nature of work, disease-related behavior, activities in males and females). It is still not clear whether the biology or behavioral/social differences drive the discrepancy found in this study.

The change of sensitivity to $PM_{2.5}$ may be related to physiological differences caused by hormonal, structural and morphological differences between males and females. Specifically, it is known that males and females differ in lung size, airway diameter, air absorption and cardiovascular response. These differences may be a significant mechanism that directly affects commuter's inhalation dose of $PM_{2.5}$, which eventually leads to different health risks[50,51]. Toxicokinetic differences in the absorption and metabolism of $PM_{2.5}$ components can affect the acceptable pollutant-related dose for males and females. Males and females have different air absorption capabilities, and the permeability of the gas-blood barrier differs by sex[52]. The inflammatory response is another important mechanism that is considered as an intermediate step in the link between air pollution and health[53]. Sexual differences in antioxidant status may have different effects on inflammatory and oxidative stress processes in males and females. However, it is not clear whether these responses biologically favor males or females.

Researches on air pollution emphasize sex-based differences when analyzing various health effects. But it is not clear whether these effects are from sex-based differences, gender-based differences, or both. The above review provides several explanations for different commuting health risks based on $PM_{2.5}$ exposure, but the direction of these effects is not always clear. Although a large number of studies have shown that females are more susceptible to outdoor air pollution, some toxicological studies suggest that males may be more susceptible due to higher doses of inhaled pollutants. Future researches should continue to investigate the sex-based differences in air pollution exposure. Collecting data on variables (activity pattern, occupations, income, socioeconomic status and etc.) that can be used to replace gender differences should be also focused. This study was able to evaluate sex-based differences but lacked information on gender or gender substitution variables.

Conclusion

Based on ECG data and PM concentration data, this study applies ROC curve to calculate the threshold of health risk during walking trips. Then a binary logit model that integrates pollutant concentration, gender, age and BMI is established. The probability of risk during trips can be effectively obtained by the proposed method. When $PM_{2.5}$ concentration is in the interval of 0–300ug/m^2, HR% gradually increases along with the rising of PM concentration. Males, youth and those with healthy BMI have better pollution resistance, as evidenced by having a higher risk threshold. Each unit increase in $PM_{2.5}$ concentration, the probability of risk will be 1.058

times higher; each unit increase in HR%, the probability of risk will be 1.114 times greater; the probability of risk for males is 76.8% of that for females; the probability of risk for youth is 67.5% of that for middle aged people; the probability of risk for normal BMI volunteers is 72.1% of that for abnormal ones. Among all indicators, $PM_{2.5}$ concentration and age impact the health risk degree more significantly. The greatest one is $PM_{2.5}$ concentration. The research is limited to the health risk of walking. While residents' travelling can be accomplished through a variety of other transportation modes, which requires further study.

References

1. Li, M. *et al.* Temporal and spatial distribution of $PM_{2.5}$ chemical composition in a coastal city of Southeast China. *Sci. Total Environ.* **605–606**, 337–346 (2017).
2. Suárez, L. *et al.* Personal exposure to particulate matter in commuters using different transport modes (bus, bicycle, car and subway) in an assigned route in downtown Santiago Chile. *Environ. Sci. Processes Impacts.* **16**, 1309–1317 (2014).
3. Yunesian, M., Rostami, R., Zarei, A., Fazlzadeh, M. & Janjani, H. Exposure to high levels of PM2. and PM10 in the metropolis of Tehran and the associated health risks during 2016–2017. *Microchemical J.* **150**, 104174 (2019).
4. Gu, Y., Deakin, E. & Long, Y. The effects of driving restrictions on travel behavior evidence from Beijing. *J. Urban Econ.* **102**, 106–122 (2017).
5. Sajjadi, S. A., Atarodi, Z., Lotfi, A. H. & Zarei, A. Levels of particulate matters in air of the Gonabad city Iran. *MethodsX* **5**, 1534–1539 (2018).
6. Duan, R. *et al.* Association between short-term exposure to fine particulate pollution and outpatient visits for ulcerative colitis in Beijing, China: A time–series study. *Ecotoxicol. Environ. Safety.* **214**, 112116 (2021).
7. Ahmadi, S. et al. Assessment of health impacts attributed to PM10 exposure during 2015–2017 in Zabol City, Iran. *Int. J. Environ. Sci. Technol.* (2021).
8. Hu, F. *et al.* The impacts of the ganglionated plexus ablation sequence on the vagal response, heart rate, and blood pressure during cardioneuroablation. *Autonomic Neurosci.* **233**, 102812 (2021).
9. Dehghani, M. H. *et al.* The data on the dispersion modeling of traffic-related PM10 and CO emissions using CALINE3; A case study in Tehran Iran. *Data Brief.* **19**, 2284–2290 (2018).
10. Amarloei, A., Fazlzadeh, M., Jafari, A. J., Zarei, A. & Mazloomi, S. Particulate matters and bioaerosols during Middle East dust storms events in Ilam. *Iran. Microchem. J.* **152**, 104280 (2020).
11. Sahu, V., Elumalai, S. P., Gautam, S., Singh, N. K. & Singh, P. Characterization of indoor settled dust and investigation of indoor air quality in different micro-environments. *Int. J. Environ. Health Res.* **28**, 419–431 (2018).
12. Gautam, S., Talatiya, A., Patel, M., Chabhadiya, K. & Pathak, P. Personal exposure to air pollutants from winter season bonfires in rural areas of Gujarat India. *Exposure Health* **12**, 89–97 (2020).
13. Gupta, A., Gautam, S., Mehta, N., Patel, M. K. & Talatiya, A. Association between changes in air quality and hospital admissions during the holi festival. *SN Appl. Sci.* **1**, 163 (2019).
14. VoPham, T., Bertrand, K. A., Tamimi, R. M., Laden, F. & Hart, J. E. Ambient PM(25) air pollution exposure and hepatocellular carcinoma incidence in the United States. *Cancer Causes Control* **29**, 563–572 (2018).
15. Wu, T., Liu, P., He, X., Xu, H. & Shen, Z. Bioavailability of heavy metals bounded to $PM_{2.5}$ in Xi'an China: Seasonal variation and health risk assessment. *Environ. Sci. Pollut. Res.* **28**, 35844–35853 (2021).
16. Tong, R., Liu, J., Wang, W. & Fang, Y. Health effects of $PM_{2.5}$ emissions from on-road vehicles during weekdays and weekends in Beijing China. *Atmospheric Environ.* **223**, 117258 (2020).
17. Qiu, Z. & Cao, H. Commuter exposure to particulate matter in urban public transportation of Xi'an, China. *J. Environ. Health Sci. Eng.* **18**, 451–462 (2020).
18. Guo, L. *et al.* Sources, environmental levels, and health risks of PM2.5-bound polycyclic aromatic hydrocarbons in energy-producing cities in northern China. *Environ. Pollut.* **272**, 116024 (2021).
19. Hung, C. *et al.* Acute particulate matter exposure is associated with disturbances in heart rate complexity in patients with prior myocardial infarction. *Sci. Total Environ.* **733**, 138842 (2020).
20. Kalantari, S., Rounds, J. D., Kan, J., Tripathi, V. & Cruz-Garza, J. G. Comparing physiological responses during cognitive tests in virtual environments vs. in identical real-world environments. *Sci. Rep.* **11**, 10227 (2021).
21. Wang, F., Chen, H., Zhu, C., Nan, S. & Li, Y. Estimating driving fatigue at a plateau area with frequent and rapid altitude change. *Sensors.* **19**, 4982 (2019).
22. Malik, M. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation* **93**, 1043–1065 (1996).
23. Sharma, S., Chandra, M. & Kota, S. H. Health effects associated with PM25: A systematic review. *Current Pollut. Rep.* **6**, 345–367 (2020).
24. Li, Y., Wang, F., Ke, H., Wang, L. & Xu, C. A driver's physiology sensor-based driving risk prediction method for lane-changing process using hidden markov model. *Sensors.* **19**, 2670 (2019).
25. Onat, B. *et al.* Determinants of exposure to ultrafine particulate matter, black carbon, and PM2.5 in common travel modes in Istanbul. *Atmospheric Environ.* **206**, 258–270 (2019).
26. Ralbovsky, N. M., Fitzgerald, G. S., McNay, E. C. & Lednev, I. K. Towards development of a novel screening method for identifying Alzheimer's disease risk: Raman spectroscopy of blood serum and machine learning. *Spectrochimica Acta Part A Molecular Biomol Spectroscopy.* **254**, 119603 (2021).
27. Huang, W. S. & Chang, Y. I. Sample size determination and treatment screening in two-stage phase II clinical trials via ROC curve. *Pharm. Stat.* **17**, 504–514 (2018).
28. Eskandari, Z. *et al.* Temporal fluctuations of PM2.5 and PM10, population exposure, and their health impacts in Dezful city Iran. *J. Environ. Health Sci. Eng.* **18**, 723–731 (2020).
29. Zhao, X., Xu, S., Rong, J. & Zhang, X. Discriminating threshold of driving fatigue based on the electroencephalography sample entropy by receiver operating characteristic curve analysis. *J. Southwest Jiaotong Univ.* **48**, 178–183 (2013).
30. Tsiroukidou, K. *et al.* Cardiorespiratory fitness predicted by fibrinogen and leptin concentrations in children with obesity and risk for diabetes: A cross-sectional study and a ROC curve analysis. *Nutrients* **13**, 674 (2021).
31. Hess, S. Conditional parameter estimates from Mixed Logit models: Distributional assumptions and a free software tool. *J. Choice Model.* **3**, 134–152 (2010).
32. World Health, O. (World Health Organization, Geneva; 2012)
33. Sun, X., Chen, H., Liu, Z. & Chen, H. Research on PM25 concentration based on dissipative structure theory: A case study of Xi'an China. *Sci. Rep.* **10**, 16435 (2020).

34. Crippa, M., Janssens-Maenhout, G., Guizzardi, D., Van Dingenen, R. & Dentener, F. Contribution and uncertainty of sectorial and regional emissions to regional and global PM2.5 health impacts. *Atmospheric Chem. Phys.* **19**, 5165–5186 (2019).
35. Zhu, C., Sun, X. & Li, Y. Forecast of urban public bicycle traffic demand by station classification. *J. Jilin Univ. Eng. Technol. Edition.* **51**, 531–540 (2021).
36. Feng, S., Gao, D., Liao, F., Zhou, F. & Wang, X. The health effects of ambient PM2.5 and potential mechanisms. *Ecotoxicol. Environ. Safety.* **128**, 67–74 (2016).
37. Shin, H. H., Fann, N., Burnett, R. T., Cohen, A. & Hubbell, B. J. Outdoor fine particles and nonfatal strokes systematic review and meta-analysis. *Epidemiology* **25**, 835–842 (2014).
38. Ren, Y. *et al.* Bioaccessibility and public health risk of heavy Metal(loid)s in the airborne particulate matter of four cities in northern China. *Chemosphere* **277**, 130312 (2021).
39. Chen, R., Jia, B., Tian, Y. & Feng, Y. Source-specific health risk assessment of PM2.5-bound heavy metals based on high time-resolved measurement in a Chinese megacity: Insights into seasonal and diurnal variations. *Ecotoxicol. Environ. Safety* **216**, 112167 (2021).
40. Sørensen, M. *et al.* Personal PM25 exposure and markers of oxidative stress in blood. *Environ. Health Perspect.* **111**, 161–166 (2003).
41. Guo, P. *et al.* Short-term effects of particle size and constituents on blood pressure in healthy young adults in Guangzhou, China. *J. Am. Heart Associat.* **10**, 019063 (2021).
42. Tsou, M.-C.M. *et al.* A community-based study on associations between PM2.5 and PM1 exposure and heart rate variability using wearable low-cost sensing devices. *Environ. Pollut.* **277**, 116761 (2021).
43. Iatsenko, D. *et al.* Evolution of cardiorespiratory interactions with age. *Philosophic. Trans. Royal Soc. A Math. Phys. Eng. Sci.* **371**, 20110622 (2013).
44. Franklin, B. *et al.* Effects of physical conditioning on cardiorespiratory function, body composition and serum lipids in relatively normal-weight and obese middle-aged women. *Int. J. Obes.* **3**, 97–109 (1979).
45. Alderete, T. L. *et al.* Longitudinal associations between ambient air pollution with insulin sensitivity, β-cell function, and adiposity in Los Angeles Latino children. *Diabetes* **66**, 1789–1796 (2017).
46. Toledo-Corral, C. M. *et al.* Effects of air pollution exposure on glucose metabolism in Los Angeles minority children. *Pediatr. Obes.* **13**, 54–62 (2018).
47. Cox, L. A. & Popken, D. A. Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States?. *Ann. Epidemiol.* **25**, 162–173 (2015).
48. Shin, H. H. *et al.* Comparison of hospitalization and mortality associated with short-term exposure to ambient ozone and PM(2.5) in Canada. *Chemosphere* **265**, 128683 (2021).
49. Bell, M. L., Zanobetti, A. & Dominici, F. Who is more affected by ozone pollution? A systematic review and meta-analysis. *Am. J. Epidemiol.* **180**, 15–28 (2014).
50. Kim, C. S. & Hu, S.-C. Total respiratory tract deposition of fine micrometer-sized particles in healthy adults: Empirical equations for sex and breathing pattern. *J. Appl. Physiol.* **101**, 401–412 (2006).
51. Johnson, K. G., Johnson, D. C., Thomas, R. J., Rastegar, V. & Visintainer, P. Cardiovascular and somatic comorbidities and sleep measures using three hypopnea criteria in mild obstructive sleep-disordered breathing: Sex, age, and body mass index differences in a retrospective sleep clinic cohort. *J. Clin. Sleep Med. JCSM Official Publ. Am. Acad. Sleep Med.* **16**, 1683–1691 (2020).
52. Pacitto, A. *et al.* The influence of lifestyle on airborne particle surface area doses received by different Western populations. *Environ. Pollut.* **232**, 113–122 (2018).
53. Ray, J. L., Fletcher, P., Burmeister, R. & Holian, A. The role of sex in particle-induced inflammation and injury. *Wiley Interdisciplin Rev. Nanomed. Nanobiotechnol.* **12**, e1589 (2020).

Acknowledgements

Author contributions

C.Z.: Writing, Methodology, Software, Data analysis. Z.F.: Data curation. L.L., X.S.: Investigation. Y.L.: Reviewing and Editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.L.

Reprints and permissions information is available at www.nature.com/reprints.

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