Supplement materials

Content:

Zhang Z, *et al. Thorax* 2019; 75:85–87. doi: 10.1136/thoraxjnl-2019-213722

Methods:

Data source

China's social health insurance schemes, including the rural new cooperative medical scheme (NCMS), the urban resident-based basic medical insurance scheme (URBMI), and the urban employee-based basic medical insurance scheme (UEBMI), more than 95% of the total population have health insurance coverage by 2011.[1 2] UEBMI covering the formal-sector employees and retirees in the urban area, which is a mandatory program administered by the China's Ministry of Human Resources and Social Security.[3]

Data were drawn from a 5% sample of the UEBMI population from January 1st, 2010 to December 31st, 2016. We used systematic sampling with a random start, in brief, we selected every *k*th record from the population of size N in such a way that a sample size of n is obtained, where K ≤ N/n. The first sample record was obtained from a table of random numbers. We have the medical information for each lung cancer patient, including date of birth, date of visit, health institute name, primary diagnosis (classified according to International Classification of Diseases, 10th edition (ICD-10))[4], and city of residence. To calculate the incident rate of lung cancer in UEBMI population, we obtained the total number of UEBMI beneficiaries for each administrative district from 2010 through 2016. Because we could not exclude the existing lung cancer patients during sampling, we excluded the patients from the nominator while calculate the lung cancer rate for 2013-2016, if the patients had filed a lung cancer claim between 2010 and 2012. As the natural history of the lung cancer, and more than half of people with lung cancer die within one year of being diagnosed,[5] we assumed the lung cancer prevalence ratio equal to the incident rate of lung cancer in this study. The study was deemed as exempt from ethical approval by the institutional review board of the Peking University, and the requirement of informed consent was waived because we only used deidentified aggregated data instead of individual data.

Incidence rate

We excluded cities without baseline air pollution exposure data, and cities without total number of URBMI beneficiaries. There were 36 cities included in the final analyses. Location of the included cities, and the total UEBMI beneficiary number were showed in **Figure S1**. We calculated the city-level incidence rate of lung cancer using the new diagnosed lung cancer cases divided by the total number of beneficiaries in this city. The incidence rate of lung cancer was estimated with stratification by area, year, and age groups. We calculated crude rates of lung cancer incidence, and the age-standardized rates (ASRs) were standardized to Six National Census in 2010 and expressed as a rate per 100,000 persons per year.

Potential confounders

We obtained the smoking prevalence from the China City Adult Tobacco Survey (CCATS), which is a nationally representative survey conducted in 2010.[6] City without smoking prevalence data was assigned a value from the closest city. Census regional variable has previously been shown to be important predictor of both air pollution distribution pattern, and distribution of cancer. [7-9] We classified China into 6 regions based on administrative area, geographical, and historical customs, including Northwest, Southwest, Central China, Eastern China, Northern China, and Northeast.

Assessment of exposure to air pollution.

Annual ambient levels of $PM_{2.5}$ from 1998 to 2016 were estimated and validated from previously published prediction models.[10 11] In brief, we estimate ground-level fine particulate matter using a neural network model that incorporated Aerosol Optical Depth (AOD) retrievals from the NASA SeaWiFS, and MISR, simulation outputs from the GEOS-Chem chemical transport model, land-use terms, meteorological measurements to predict annual concentrations of $PM_{2.5}$ at unmonitored locations. Cross-validation indicated good agreement between predicted values and ground-measured $PM_{2.5}$ from the China Environmental Monitoring Center

Zhang Z, *et al. Thorax* 2019; 75:85–87. doi: 10.1136/thoraxjnl-2019-213722

(January 2013 to December 2017) or data from the U.S. consulate sites (including Beijing, Shanghai, Guangzhou, Shenyang, and Chengdu). To improve the accuracy of exposure assessment and capture the smaller scale variability of the exposure, we calculated the annual $PM_{2.5}$ concentration based on data from 10 locations which equally distributed in each city. Because the appropriate exposure period is unknown, we assumed the development of lung cancer to be more chronic than short-term. Thus, 1-, 3-, 5-, and 10-year cumulative average $PM_{2.5}$ concentration was considered.

Statistical analysis

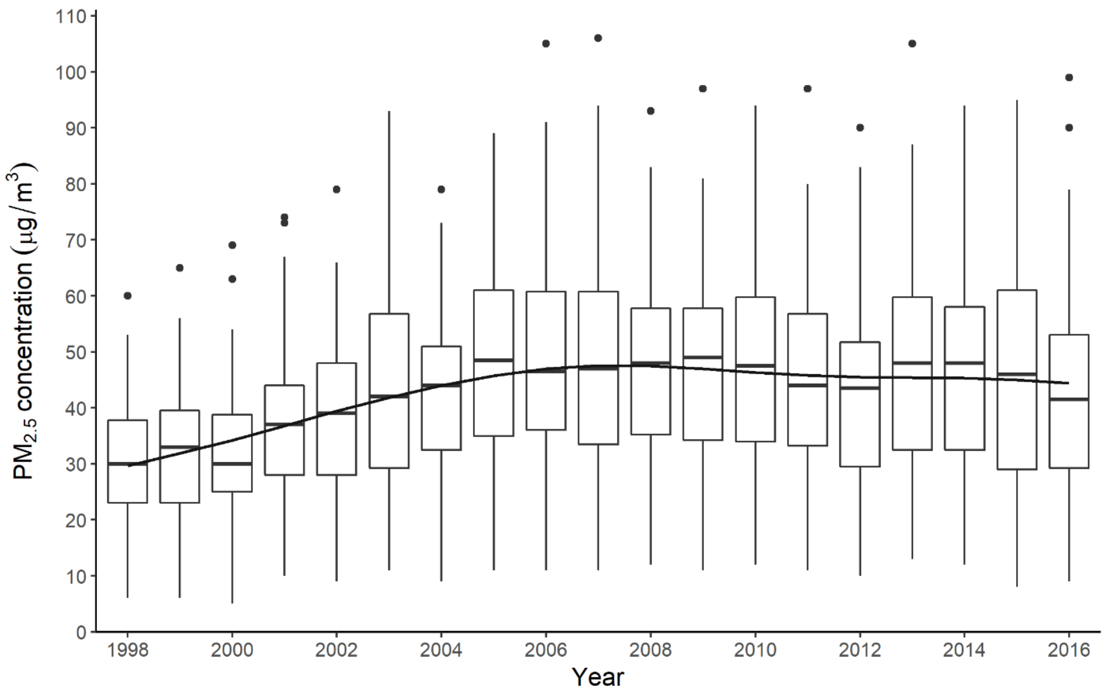
The association of lung cancer incident rate with long-term $PM_{2.5}$ exposure was the primary interest. As the lung cancer rate was calculated with an aggregated data, we used the weighted logistic regression model to estimate risk ratio (RRs) and confidence intervals (CIs) for incidence rates of lung cancer with 10 $\mu g/m^3$ increase in $PM_{2.5}$ exposure, with lung cancer incident rate using as the weight variable. We investigated potential effects modification by age group (younger than 30 years, 30-44 years, 45-59 years, 60-74 years, and 75 years or older), and low/high exposure (cities with annual mean $PM_{2.5}$ concentration under 35 $\mu g/m^3$ were classified as low exposure group, according to the WHO's air quality guideline for Interim target (IT) 1 (IT1: $PM_{2.5}$ <35 $\mu g/m^3$) [12]).

Population attributive fraction (PAFs) were estimated using the following equation: PAF= 100 × k (adjusted RR − 1) / (k [adjusted RR − 1] + 1), for which k indicates the prevalence of URBMI beneficiaries lived in city with higher $PM_{2.5}$ concentration level (annual average concentration ≥ 35 $\mu g/m^3$). The PAFs were calculated from RR of lung cancer for people exposure to higher pollution concentration level (≥ 35 $\mu g/m^3$) with people exposure to lower pollution concentration level (<35 $\mu g/m^3$). Confidence intervals (CIs) for PAFs were obtained by bootstrapping (1,000 iterations).

To evaluate the non-linear concentration-response relationship between $PM_{2.5}$ air pollution and lung cancer, we modelled $PM_{2.5}$ air pollution using restricted cubic splines with knots at the 5^{th}, 35th, 65th, and 95th percentiles of the distribution of $PM_{2.5}$ concentrations. The reference exposure level was set at the 10th percentile of the distribution of $PM_{2.5}$ concentrations. Statistical analyses were conducted using STATA version 15.0 (Stata Corp LP, College Station, TX) and R version 3.4.1 (R Foundation for Statistical Computing).

Zhang Z, *et al. Thorax* 2019; 75:85–87. doi: 10.1136/thoraxjnl-2019-213722

Figure S1. Average annual $PM_{2.5}$ concentrations in China, 1998-2016.

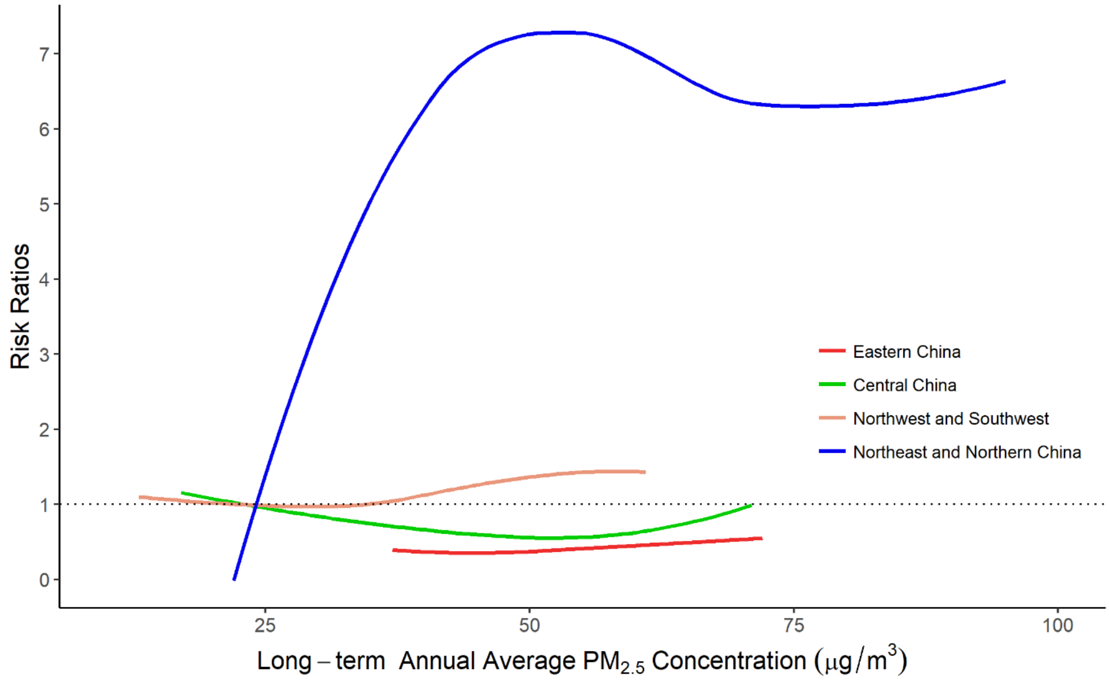


Zhang Z, *et al. Thorax* 2019; 75:85–87. doi: 10.1136/thoraxjnl-2019-213722

99 **Figure S2.** Distribution of cities and the total number of beneficiaries

100

Figure S3. Risk ratios for lung cancer associated with a 10 µg/m³ increase in $PM_{2.5}$ concentration, by difference exposure periods.



Risk ratios were adjusted for calendar time, region, smoking rate. The dose-response curve was calculated using restricted cubic splines with knots at the 5th, 35th, 65th, and 95th percentiles of the distribution of $PM_{2.5}$ concentrations. The reference exposure level was set at the 10th percentile of the distribution of 1-year $PM_{2.5}$ concentration (23.0 µg/m³). Eastern China includes Shanghai, Suzhou, Zhejiang, Anhui, Fujian, Jiangxi, and Shandong province; Central China includes Henan, Hubei, Hunan, Guangdong, Guangxi, and Hainan province; Northwest includes Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang province; Southwest includes Chongqing, Guizhou, Yunnan, and Xizang province; Northeast includes Liaoning, Jilin, and Heilongjiang province; Northern China includes Beijing, Tianjin, Hebei, Shanxi, and Inner Mongolia.

Zhang Z, *et al. Thorax* 2019; 75:85–87. doi: 10.1136/thoraxjnl-2019-213722

Table S1. Lung cancer rate in 5% sample of Urban Basic Medical Insurance population (by age groups).

	Crude rates ($1/10^5$)	ASR ($1/10^5$)
Age groups		
<30	192	14.8
30-44	416	39
45-59	224	141
60-74	5,138	153
75+	5,000	56.4

ASR: Age standardized rate.

Zhang Z, *et al. Thorax* 2019; 75:85–87. doi: 10.1136/thoraxjnl-2019-213722

Table S2. City specific incidence rate of lung cancer, smoking rate, and pollution levels.

City	Incidence rate	Smoking Rate	1-year $PM_{2.5}$	3-year $PM_{2.5}$	5-year $PM_{2.5}$	10-year $PM_{2.5}$
Shanghai	57.7	20.5	53	50.6	51.2	52.9
Dongguan	85.1	18.6	39	38.4	38.9	41.7
Lanzhou	140	22	35.5	35.3	35.9	39.2
Baotou	35.5	22.3	28.3	26.9	26.1	26.5
Beijing	188	22.3	76	75.8	76.3	77.6
Nanjing	26	20.5	63	61.1	61.4	61.9
Nanchang	120	22.5	45	44.7	45.5	46.9
Nantong	89.7	20.5	55.7	56.3	56.6	58.2
Xiamen	78.5	18.6	38.2	38.2	38.6	40.1
Hefei	108	22.5	60.7	57.6	57.6	58.3
Wuzhong	106	22	23	25	23.4	24.9
Huhehaote	70.7	22.3	24.5	24.2	24.3	24.4
Haerbin	170	18.8	55.5	49.4	47.8	45.4
Datong	111	22	40	38.8	38.1	38.4
Dalian	264	24.5	40.7	38	39.1	40.7
Tianjin	217	22.1	76.8	75.2	75	75.2
Taiyuan	124	23	50.8	51.4	52.4	55.8
Guangzhou	69.8	18.6	43.2	43.2	44.2	46.2
Chengdu	137	22	55.2	57.6	57.4	59.2
Lasa	10.8	21.3	13	12.3	12.2	11.7
Kunming	113	22	17	17.8	18.2	18.9
Hangzhou	210	20.5	61.2	60.4	60.2	61.6
Liuzhou	177	18.6	43	43.8	44.2	44.8
Wuhan	59.3	23	64.3	60.1	60.3	61.6
Shenyang	219	24.5	46	47.2	50.2	47.4
Haikou	211	18.6	20.5	20.4	20.7	21.4
Shenzhen	17.7	18.6	32	32.7	33.5	34.5
Shijiazhuang	120	22.1	96	96.3	95.8	96.1
Fuzhou	116	18.6	40.7	39.4	39.3	40.3
Qinhuangdao	69.8	22.1	52	54.5	55.9	55.7
Jingzhou	94.2	23	58.5	58.2	58	57.8
Xining	110	22	28.5	27.8	28.3	28
Guiyang	113	22	31	32.2	32.9	33.6
Zunyi	101	22	40.3	42.3	43.4	44.8
Zhengzhou	77.4	21.3	67.7	72.6	71.5	73.8
Changchun	195	23.5	52.8	48.5	48	45.2

Reference

1. Meng Q, Fang H, Liu X, et al. Consolidating the social health insurance schemes in China: towards an equitable and efficient health system. 2015;386(10002):1484-92.

2. Statistics NBo. China Labor Statistical Yearbook. 2015

3. Barber SL, Yao L. Health insurance systems in China: a briefing note. *World health report* 2010;37

4. World Health Organization (2007) International Statistical Classification of Diseases and Related Health Problems tRv, 2nd ed Geneva: WHO.

5. Noone AM, Howlader N, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2015, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2015/, based on November 2017 SEER data submission, posted to the SEER web site, April 2018. 2017

6. X L. Report of China City Adult Tobacco Survey 2013–14. CDC Foundation, Atlanta, Georgia, USA. 2015

7. Yanosky JD, Paciorek CJ, Schwartz J, et al. Spatio-temporal modeling of chronic PM10 exposure for the Nurses' Health Study. 2008;42(18):4047-62.

8. Hart JE, Yanosky JD, Puett RC, et al. Spatial modeling of PM10 and NO2 in the continental United States, 1985–2000. 2009;117(11):1690-96.

9. Zhou MG, Wang XF, Hu JP, et al. [Geographical distribution of cancer mortality in China, 2004-2005]. *Zhonghua Yu Fang Yi Xue Za Zhi* 2010;44(4):303-8. [published Online First: 2010/07/27]

10. Zhang Z, Wang J, Hart JE, et al. National scale spatiotemporal land-use regression model for PM2. 5, PM10 and NO2 concentration in China. *Atmospheric Environment* 2018;192:48-54.

11. Boys B, Martin R, Van Donkelaar A, et al. Fifteen-year global time series of satellite-derived fine particulate matter. *Environmental science & technology* 2014;48(19):11109-18.

12. WHO. Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide. Global update 2005. Summary of risk assessment. 2006

Zhang Z, *et al. Thorax* 2019; 75:85–87. doi: 10.1136/thoraxjnl-2019-213722