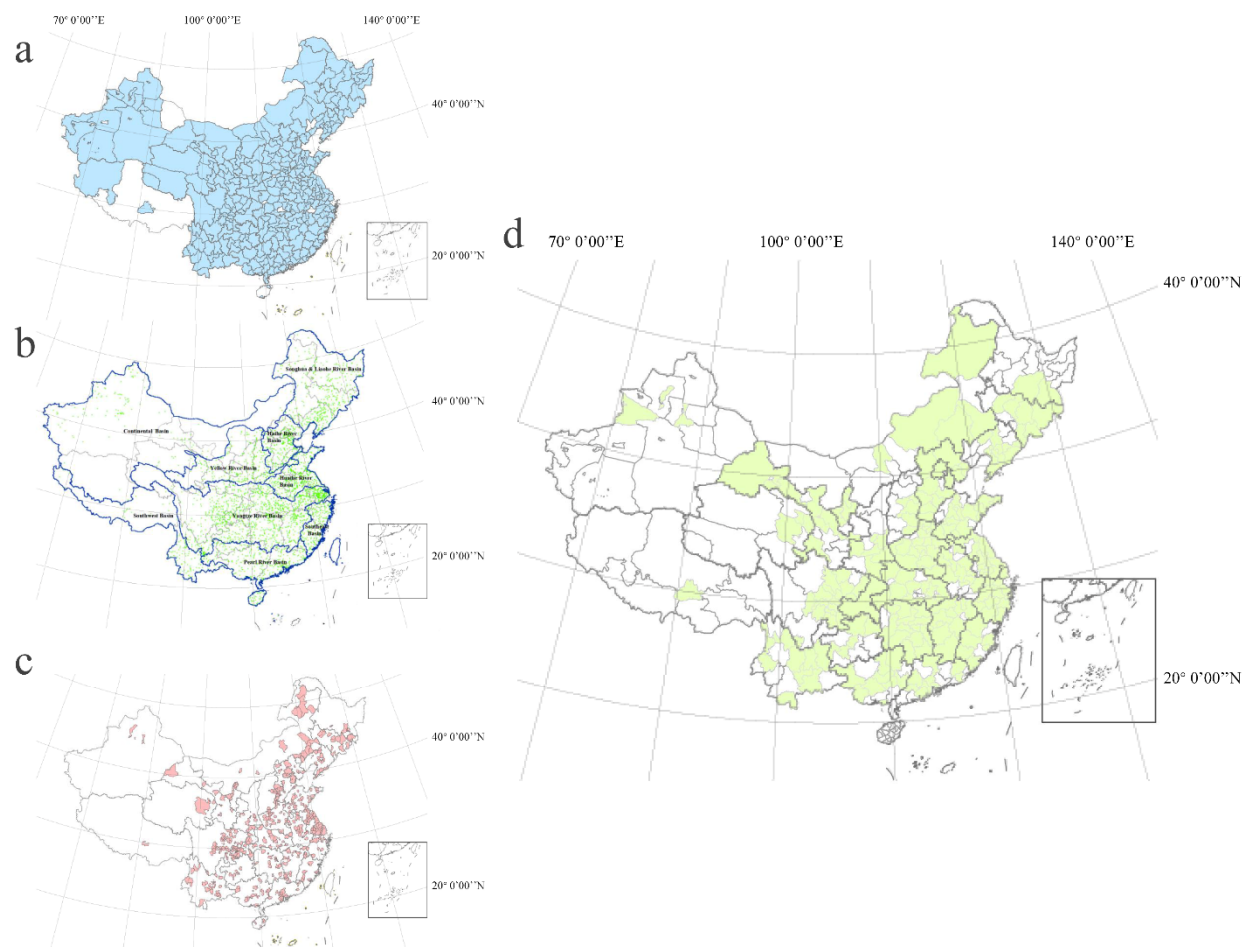
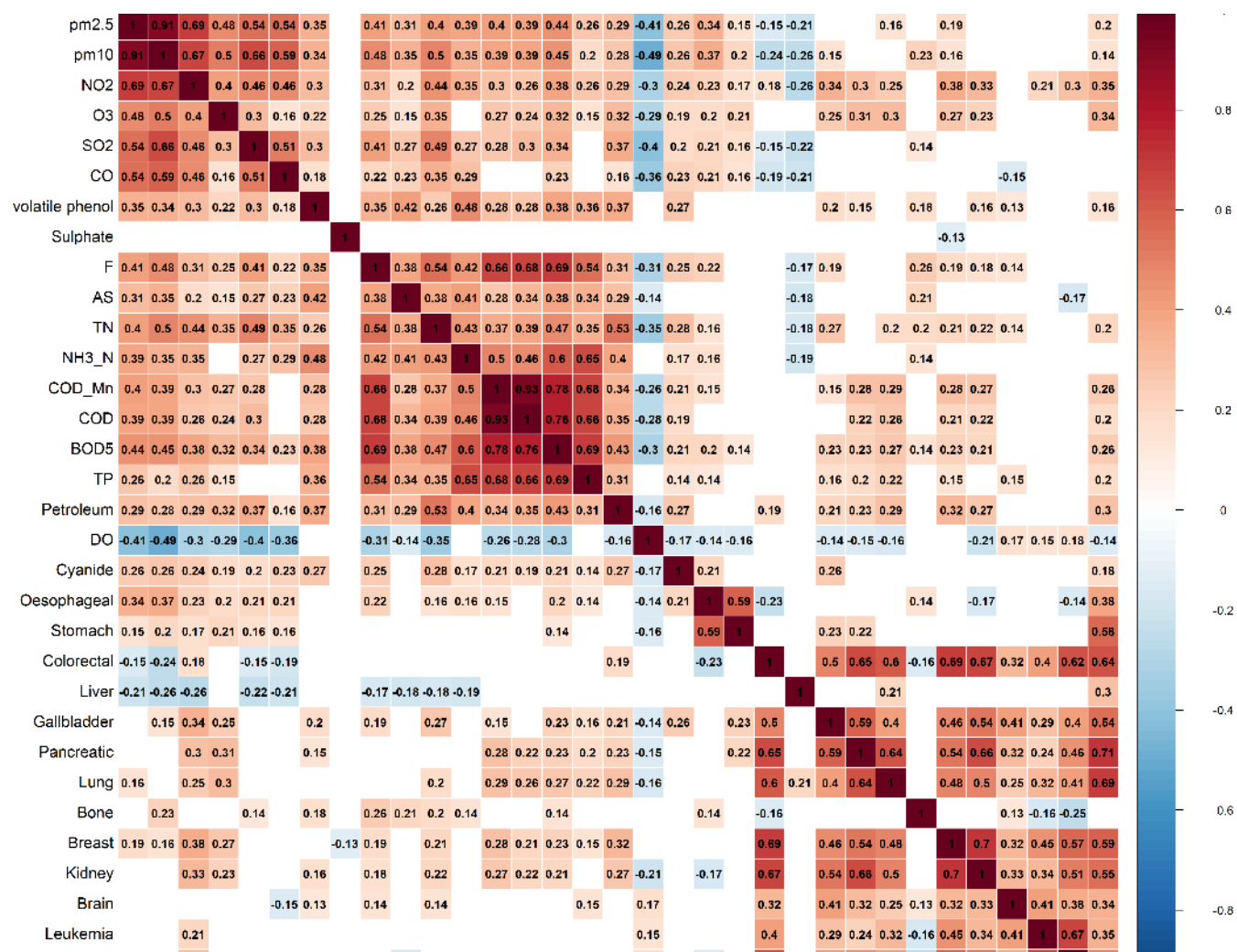


Supplementary Information



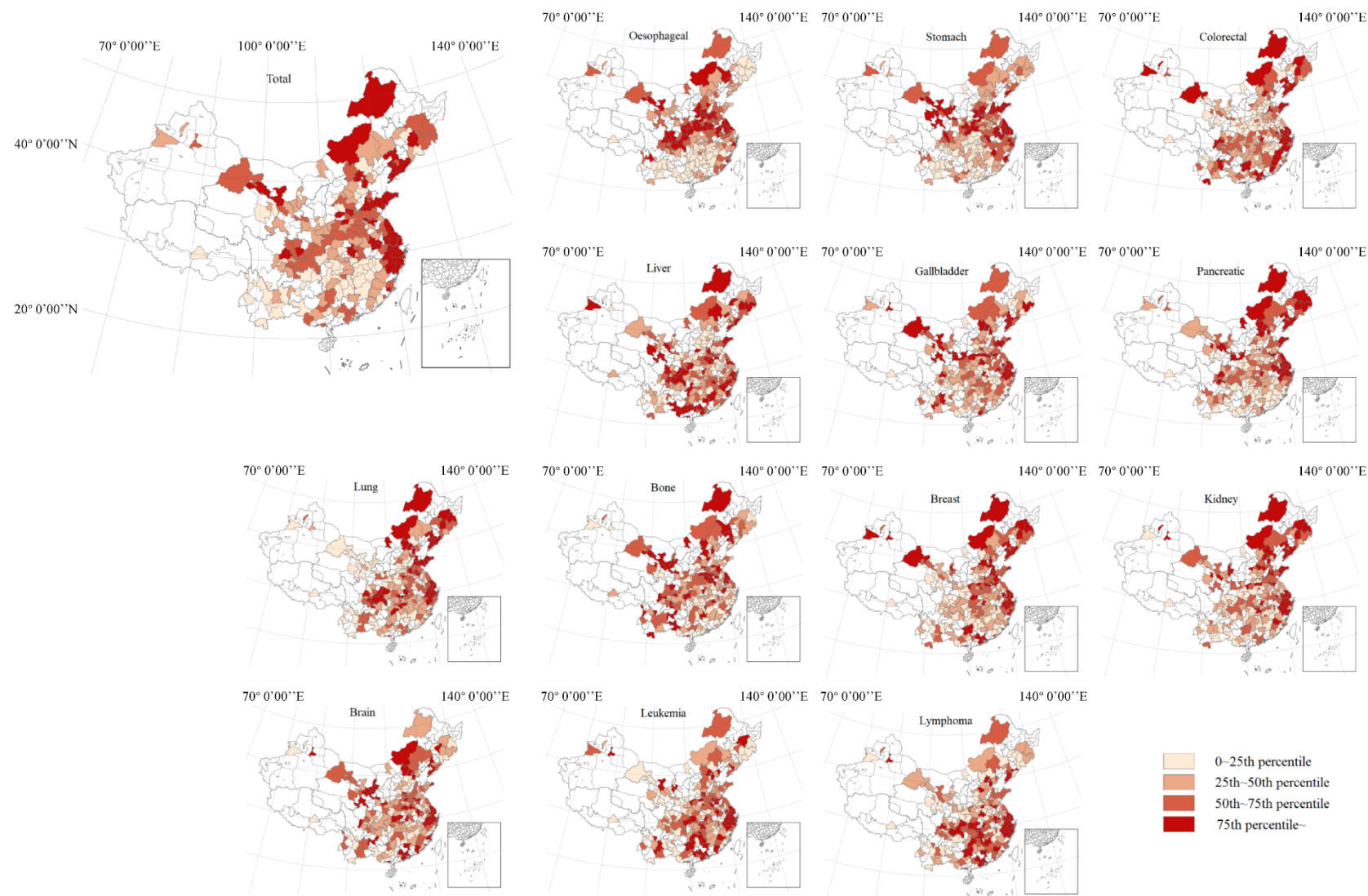
Supplementary Information Fig. 1 Spatial distribution of 331 prefecture-level areas with air monitoring sites (a), 3632 water quality monitoring sections (b), 487 CRIs (c), and 219 basic analytic units (d).

The base map was obtained from the China Resource and Environment Science and Data Center.



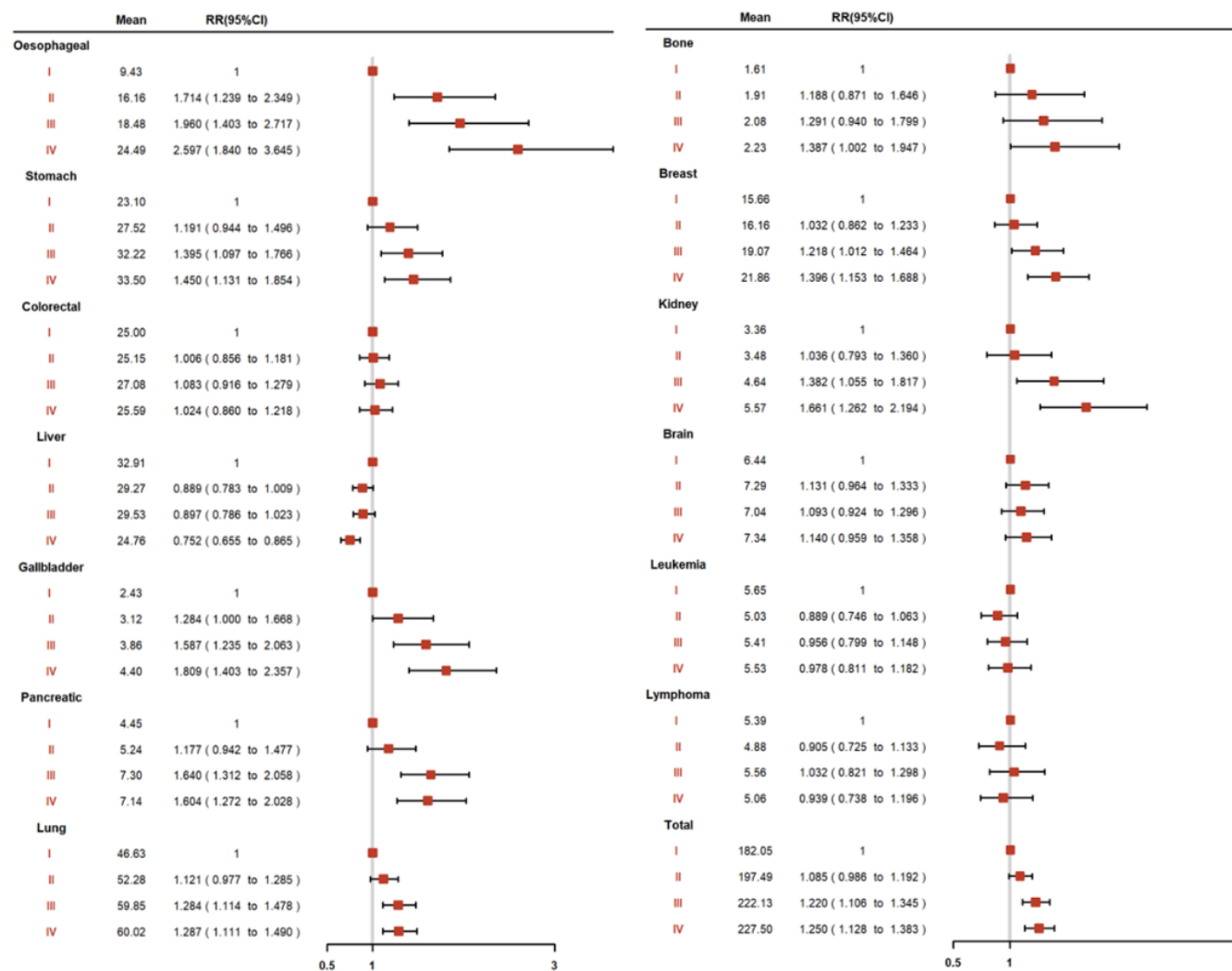
Supplementary Information Fig. 2 Spearman correlation coefficient diagram between six air pollutants, 13 surface water pollutants and incidence of 13 types of cancer.

Source data are provided as a Source Data file.



Supplementary Information Fig. 3 Spatial distributions of incidence (1/100, 000) of 13 types of cancer. P25, P50, P75 denote the 25th, 50th, 75th percentiles of the incidence of each cancer type, respectively.

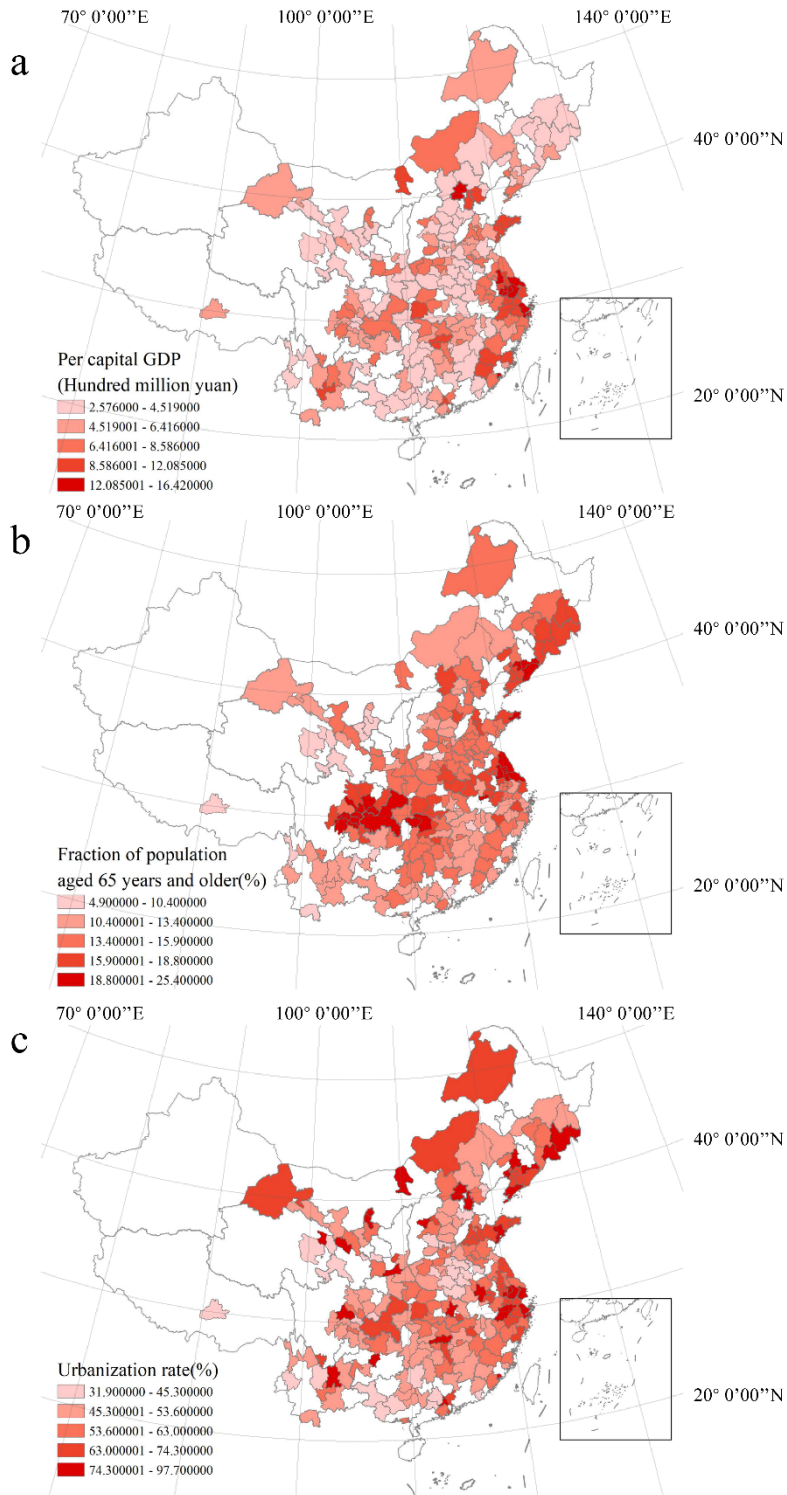
The base map was obtained from the China Resource and Environment Science and Data Center.



Supplementary Information Fig. 4 Sensitivity analysis results: estimated rate ratios and 95% confidence intervals for incidence of 13 types of cancer in relation to an alternative grouping scheme for defining co-pollution grades.

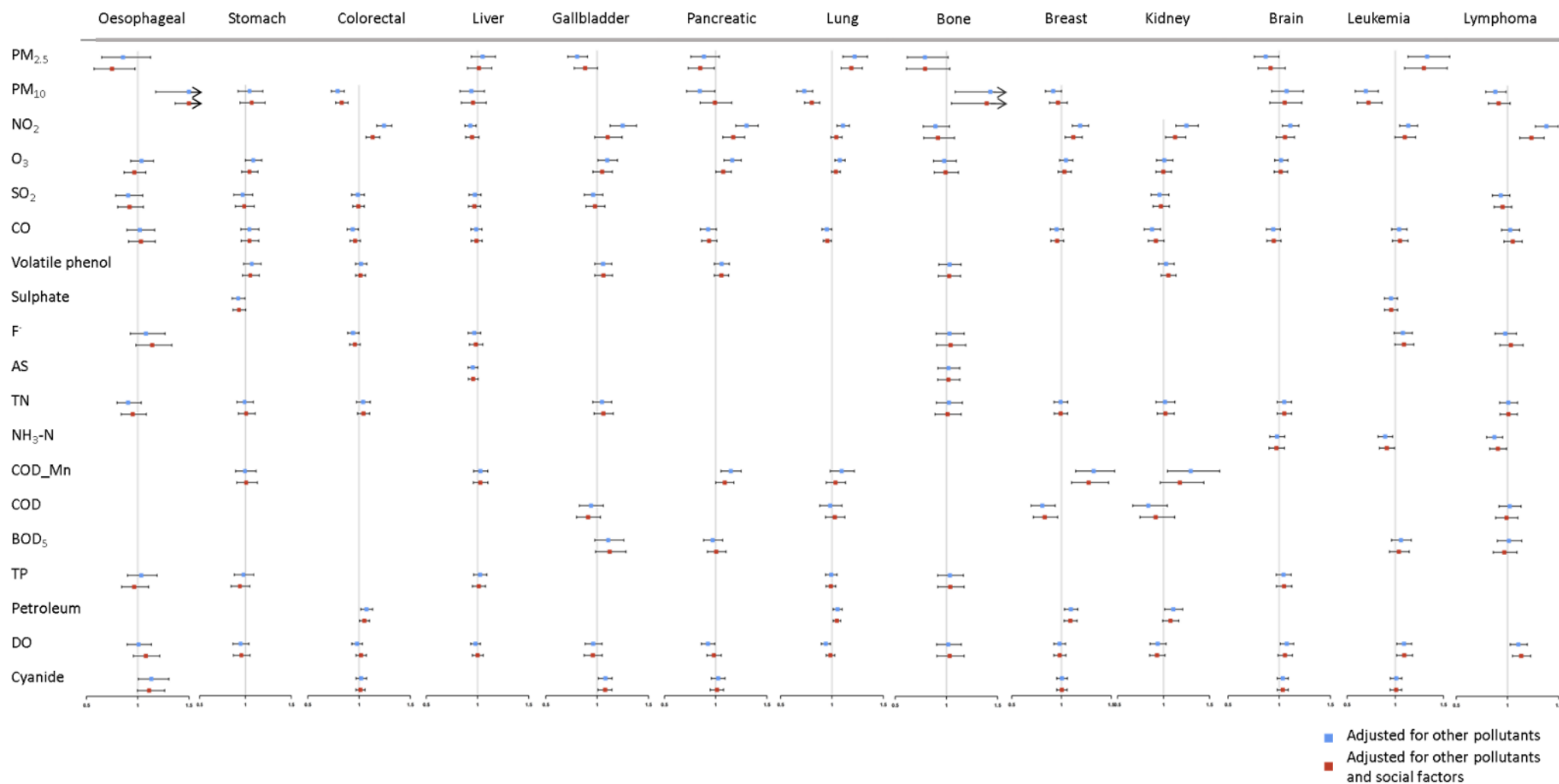
For air, units with 0–1, 2–3, and 4–6 “H” pollutants were defined as 1, 2, and 3 (low- to high-level pollution), respectively. For water, units with 0–1, 2–6, and 7–13 “H” pollutants were defined as 1, 2, and 3 (low- to high-level pollution). There were 32, 77, 61, and 49 units in Grade I, II, III, and IV, respectively.

Source data are provided as a Source Data file.



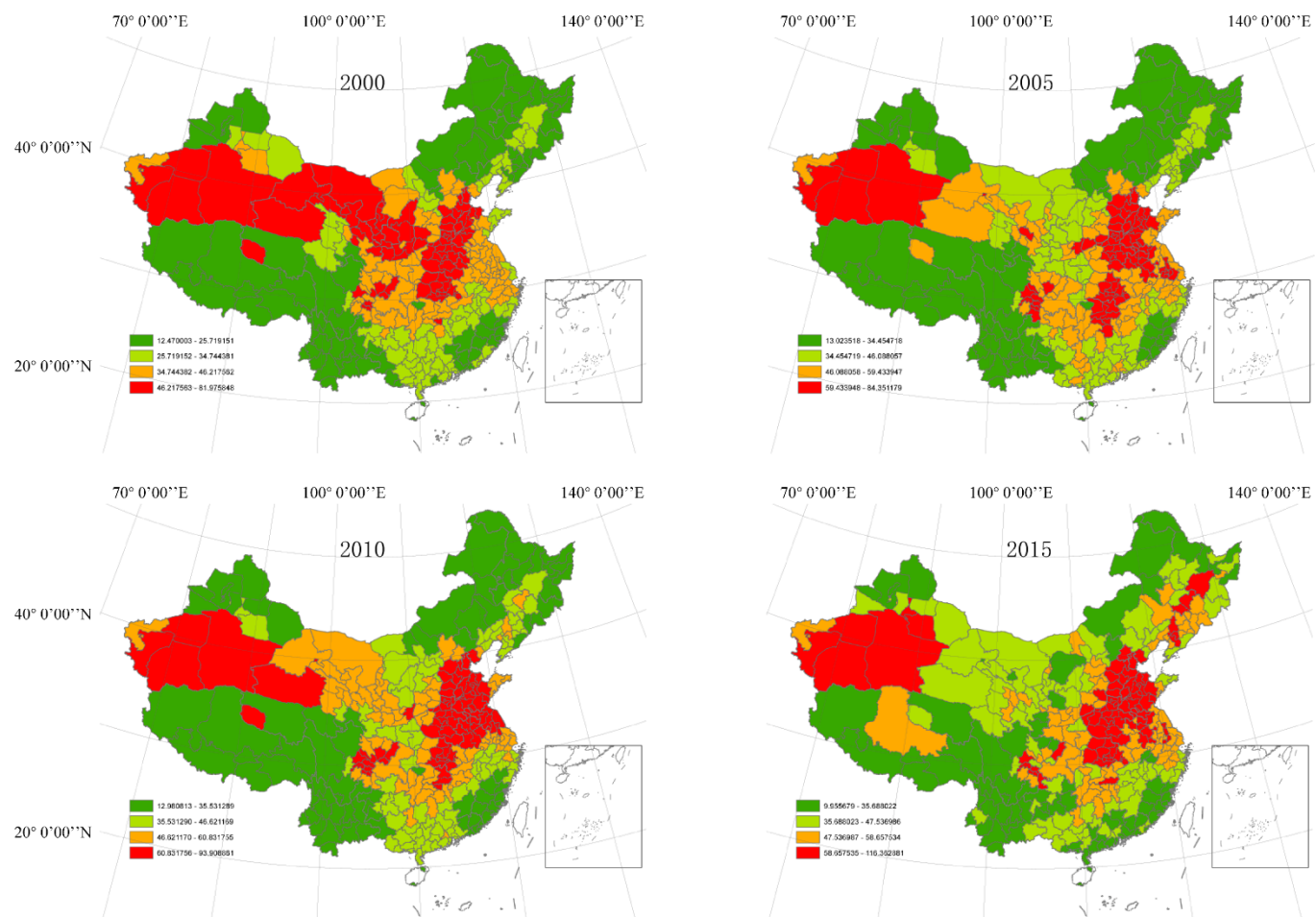
Supplementary Information Fig. 5 Spatial distributions of social factors, per capital gross domestic product (a), fraction of population aged 65 years and older (b), and urbanisation rate (c).

The base map was obtained from the China Resource and Environment Science and Data Center.



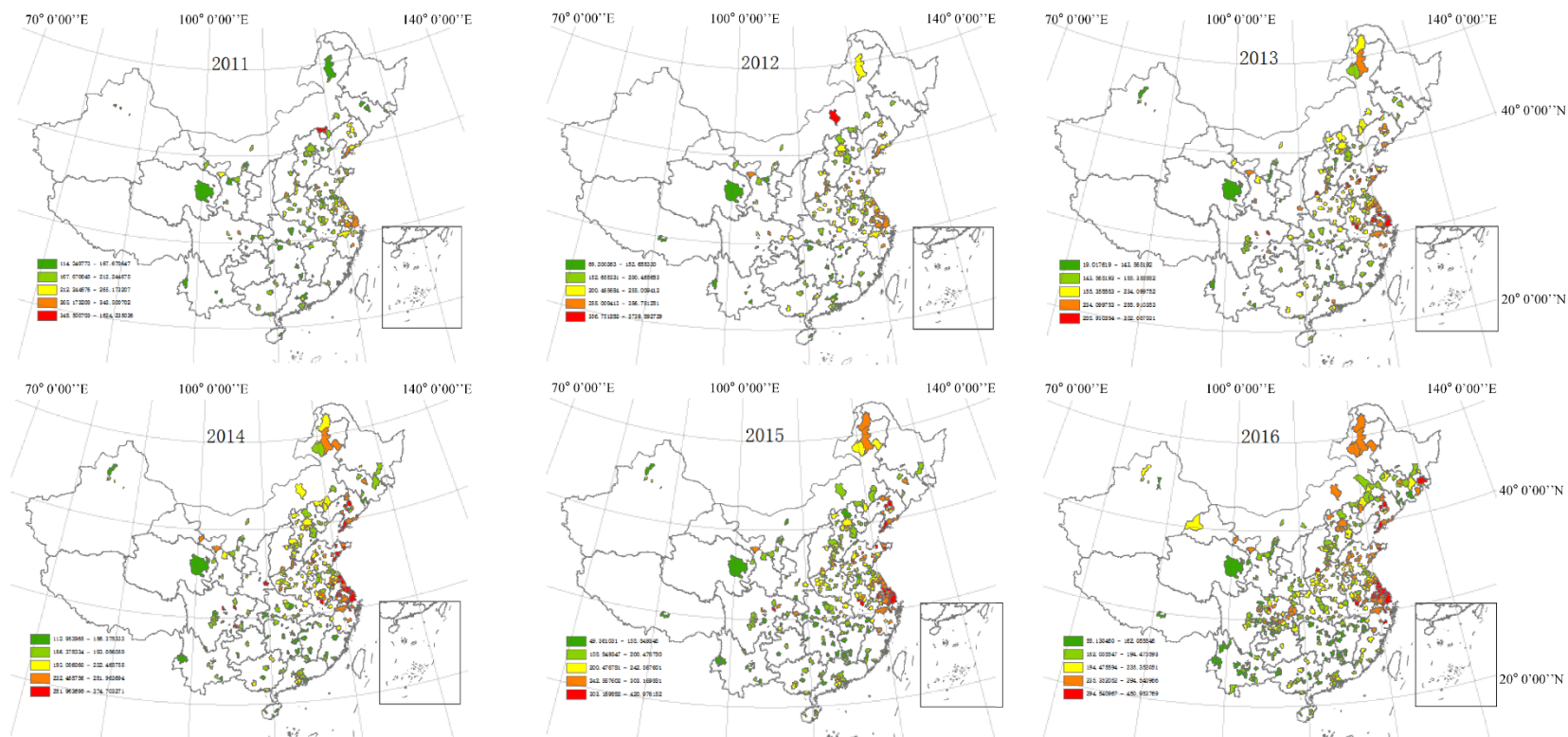
Supplementary Information Fig. 6 Estimated rate ratios and 95% confidence intervals of key pollutants on each cancer, before and after adjusting for social factors.

Source data are provided as a Source Data file.



Supplementary Information Fig. 7 Spatial distribution of PM_{2.5} concentration ($\mu\text{g}/\text{m}^3$) in 2000, 2005, 2010, and 2015. Units were categorized by national yearly quartile levels.

The base map was obtained from the China Resource and Environment Science and Data Center.



Supplementary Information Fig. 8 Spatial distribution of incidence (1/100, 000) of total cancer in 2011-2016.

The base map was obtained from the China Resource and Environment Science and Data Center.

Supplementary Information Table 1. Frequency of air and water pollutant monitoring site/section and cancer registry institute (CRI) in 30 provinces.

Province	Prefecture-level units/ air pollutant monitoring site (n)	Water pollutant monitoring section (n)	CRI (n)
Beijing	1	33	2
Tianjin	1	44	2
Hebei	10	83	27
Shanxi	5	20	7
Inner Mongolia	5	46	15
Liaoning	9	85	14
Jilin	6	69	15
Heilongjiang	2	12	8
Shanghai	1	41	1
Jiangsu	11	196	42
Zhejiang	11	108	14
Anhui	11	130	22
Fujian	7	52	11
Jiangxi	10	115	24
Shandong	12	108	23
Henan	16	112	32
Hubei	11	150	15
Hunan	13	88	22
Guangdong	12	75	15
Guangxi	9	52	18
Chongqing	1	82	24
Sichuan	17	134	48
Guizhou	5	38	9
Yunnan	9	97	24
Tibet	1	3	1
Shaanxi	6	31	17
Gansu	7	17	9
Qinghai	3	13	4
Ningxia	4	13	8
Xinjiang	3	6	4
Total	219	2053	477

Supplementary Information Table 2. Basic descriptive characteristic of six air pollutants, 13 surface water pollutants and incidence of 13 types of cancer.

	Mean (SD)	Min	P25	P50	P75	Max
Air pollutant concentration						
PM _{2.5} (µg/m ³)	53.08 (17.37)	15.77	40.93	51.68	62.19	106.12
PM ₁₀ (µg/m ³)	89.61 (29.90)	31.01	67.89	85.62	106.13	175.14
NO ₂ (µg/m ³)	31.71 (10.49)	7.68	23.42	30.63	39.80	61.40
O ₃ (µg/m ³)	106.88 (20.98)	45.20	92.44	106.80	121.82	155.49
CO (mg/m ³)	1.10 (0.34)	0.46	0.88	1.03	1.30	2.15
SO ₂ (µg/m ³)	26.69 (13.86)	6.35	17.11	22.93	33.13	79.38
Water pollutant concentration						
Volatile phenol (mg/L)	0.00042 (0.00025)	0.0002	0.00024	0.00035	0.00053	0.00199
Sulphate (mg/L)	0.00353 (0.00313)	0.002	0.002	0.00244	0.00382	0.02769
F ⁻ (mg/L)	0.34992 (0.20142)	0.0664	0.19198	0.29112	0.44936	1.201
AS (mg/L)	0.02648 (0.00862)	0.02	0.02	0.02339	0.02928	0.07111
TN (mg/L)	2.95511 (1.84879)	0.215	1.68745	2.30218	3.80191	11.21604
NH ₃ -N (mg/L)	0.24772 (0.16928)	0.02208	0.13233	0.20906	0.3037	1.16536
COD_Mn (mg/L)	3.13117 (1.40391)	0.84861	2.03333	2.88	4.03167	10.97859
COD (mg/L)	12.23463 (4.51246)	3.375	8.66667	11.6963	15.06384	35.07117
BOD ₅ (mg/L)	1.74883 (0.72214)	0.43889	1.22867	1.61409	2.14074	3.83093
TP (mg/L)	0.07818 (0.03590)	0.01063	0.05346	0.07326	0.09907	0.24155
Petroleum (mg/L)	0.00869 (0.00420)	0.0045	0.00555	0.00704	0.01031	0.02592
DO (mg/L)	8.78993 (0.93949)	5.96875	8.28667	8.77967	9.40562	12.02847
Cyanide (mg/L)	0.00168 (0.00075)	0.0005	0.00128	0.00181	0.002	0.00941
Cancer incidence (per 100, 000)						
Oesophageal	17.69 (15.95)	0	6.86	12.07	22.8	82.33
Stomach	29.52 (18.36)	5.80	16.13	25.01	38.14	123.18
Colorectal	25.77 (10.40)	1.90	18.31	23.11	32.28	74.11
Liver	28.87 (10.06)	5.16	22.57	27.42	32.05	86.21
Gallbladder	3.52 (1.99)	0	1.94	3.06	4.65	9.67
Pancreatic	6.13 (3.39)	0.48	3.68	5.16	7.46	22.00
Lung	55.30 (18.84)	9.51	43.28	52.83	65.55	151.76
Bone	1.98 (0.96)	0.22	1.39	1.81	2.38	6.23
Breast	18.17 (8.03)	2.01	12.51	16.72	23.06	46.88
Kidney	4.25 (2.87)	0	2.39	3.58	5.10	15.57
Brain	7.11 (2.74)	1.28	5.17	6.77	8.68	16.82
Leukemia	5.34 (2.29)	0	3.87	5.32	6.82	12.27
Lymphoma	5.19 (2.79)	0	3.36	4.94	6.78	15.14

Note: P25, P50, P75 denote the 25th, 50th, 75th percentiles.

Supplementary Information Table 3. Results of global Moran's *I* of six air pollutants and 13 surface water pollutants.

Pollutant	Moran's <i>I</i>	Z score	P value
PM _{2.5} (µg/m ³)	0.17	15.02	<0.001
PM ₁₀ (µg/m ³)	0.35	29.49	<0.001
NO ₂ (µg/m ³)	0.20	17.42	<0.001
O ₃ (µg/m ³)	0.16	13.92	<0.001
CO(mg/m ³)	0.04	4.24	<0.001
SO ₂ (µg/m ³)	0.04	3.86	<0.001
Volatile phenol (mg/L)	0.13	11.29	<0.001
Sulphate (mg/L)	0.06	5.18	<0.001
F ⁻ (mg/L)	0.37	31.83	<0.001
AS (mg/L)	0.05	4.65	<0.001
TN (mg/L)	0.36	30.22	<0.001
NH ₃ -N (mg/L)	0.12	10.56	<0.001
COD_Mn (mg/L)	0.24	20.69	<0.001
COD (mg/L)	0.25	21.47	<0.001
BOD ₅ (mg/L)	0.31	26.18	<0.001
TP (mg/L)	0.08	7.46	<0.001
Petroleum (mg/L)	0.20	17.31	<0.001
DO (mg/L)	0.28	24.10	<0.001
Cyanide (mg/L)	0.11	9.51	<0.001

Note: The *p*-value was calculated using the two-sided Z test.

Supplementary Information Table 4. The attributable proportion of each cancer type in areas with co-pollution Grades I–IV (%).

Co-pollution grade	Pollutants	Oesophageal	Stomach	Colorectal	Liver	Gallbladder	Pancreatic	Lung	Bone	Breast	Kidney	Brain	Leukemia	Lymphoma
Grade I	Overall	3.35	0	0	0	0	0	4.01	1.82	0	0	0	12.02	8.78
	PM _{2.5}	0	0	0	0	0	0	4.01	0	0	0	0	6.01	0
	PM ₁₀	3.35	0	0	0	0	0	0	1.82	0	0	0	0	0
	NO ₂	0	0	0	0	0	0	0	0	0	0	0	0	0
	O ₃	0	0	0	0	0	0	0	0	0	0	0	0	0
	COD_Mn	0	0	0	0	0	0	0	0	0	0	0	0	0
	Petroleum	0	0	0	0	0	0	0	0	0	0	0	0	0
	DO	0	0	0	0	0	0	0	0	0	0	0	6.39	8.78
	Cyanide	0	0	0	0	0	0	0	0	0	0	0	0	0
Grade II		Oesophageal	Stomach	Colorectal	Liver	Gallbladder	Pancreatic	Lung	Bone	Breast	Kidney	Brain	Leukemia	Lymphoma
	Overall	14.17	0	0	0	0	0	4.49	7.77	0	0	0	10.35	5.40
	PM _{2.5}	0	0	0	0	0	0	4.49	0	0	0	0	6.7	0
	PM ₁₀	14.17	0	0	0	0	0	0	7.77	0	0	0	0	0
	NO ₂	0	0	0	0	0	0	0	0	0	0	0	0	0
	O ₃	0	0	0	0	0	0	0	0	0	0	0	0	0
	COD_Mn	0	0	0	0	0	0	0	0	0	0	0	0	0
	Petroleum	0	0	0	0	0	0	0	0	0	0	0	0	0
	DO	0	0	0	0	0	0	0	0	0	0	0	3.91	5.40
Grade III		Oesophageal	Stomach	Colorectal	Liver	Gallbladder	Pancreatic	Lung	Bone	Breast	Kidney	Brain	Leukemia	Lymphoma
	Overall	21.09	0	7.99	0	0.86	11.59	5.63	11.52	10.49	7.63	0	12.69	18.71
	PM _{2.5}	0	0	0	0	0	0	5.63	0	0	0	0	8.35	0
	PM ₁₀	20.14	0	0	0	0	0	0	11.52	0	0	0	0	0
	NO ₂	0	0	7.99	0	0	10.06	0	0	7.27	7.63	0	0	13.04
	O ₃	0	0	0	0	0	0.45	0	0	0	0	0	0	0
	COD_Mn	0	0	0	0	0	1.26	0	0	3.47	0	0	0	0
	Petroleum	0	0	0	0	0	0	0	0	0	0	0	0	0

Grade IV	DO	0	0	0	0	0	0	0	0	0	0	0	4.73	6.51
	Cyanide	1.19	0	0	0	0.86	0	0	0	0	0	0	0	0
		Oesophageal	Stomach	Colorectal	Liver	Gallbladder	Pancreatic	Lung	Bone	Breast	Kidney	Brain	Leukemia	Lymphoma
	Overall	25.53	0	10.45	0	2.63	20.33	7.70	13.14	24.75	8.03	0	8.56	13.79
	PM _{2.5}	0	0	0	0	0	0	5.75	0	0	0	0	8.53	0
	PM ₁₀	22.71	0	0	0	0	0	0	13.14	0	0	0	0	0
	NO ₂	0	0	8.4	0	0	10.57	0	0	7.65	8.03	0	0	13.71
	O ₃	0	0	0	0	0	5.46	0	0	0	0	0	0	0
	COD_Mn	0	0	0	0	0	5.77	0	0	15.32	0	0	0	0
	Petroleum	0	0	2.24	0	0	0	2.07	0	3.79	0	0	0	0
	DO	0	0	0	0	0	0	0	0	0	0	0	0.03	0.09
	Cyanide	3.64	0	0	0	2.63	0	0	0	0	0	0	0	0

PM_{2.5}, particulate matter ≤ 2.5 ; PM₁₀, particulate matter ≤ 10 ; NO₂, nitrogen dioxide; O₃, ozone; DO, dissolved oxygen.

Supplementary Information Table 5. Comparison of characteristics of included areas and not-included areas.

	Areas included	Areas not included	P value
	Mean (SD)	Mean (SD)	
Air pollutant concentration			
PM _{2.5} (µg/m ³)	53.08(17.37)	44.01(19.21)	<0.001
PM ₁₀ (µg/m ³)	89.61(29.9)	81.53(44.03)	0.0851
NO ₂ (µg/m ³)	31.71(10.49)	25.01(10.34)	<0.001
O ₃ (µg/m ³)	106.88(20.98)	103.09(20.68)	0.119
CO (mg/m ³)	1.10(0.34)	1.05(0.42)	0.303
SO ₂ (µg/m ³)	26.69(13.86)	24.37(15.13)	0.178
Water pollutant concentration			
Volatile phenol (mg/L)	0.00021(0.00043)	0.00019(0.0003)	<.001
Sulphate (mg/L)	0.00176(0.0025)	0.00146(0.00191)	<.001
F ⁻ (mg/L)	0.18077(0.19543)	0.18544(0.25052)	<.001
AS (mg/L)	0.0121(0.00784)	0.01276(0.01246)	<.001
TN (mg/L)	2.61711(2.40154)	2.23582(2.14091)	<.001
NH ₃ -N (mg/L)	0.22809(0.27361)	0.2231(0.34182)	<.001
COD_Mn (mg/L)	3.08613(1.67371)	2.84099(1.85485)	<.001
COD (mg/L)	8.70467(6.70761)	8.256(6.83863)	0.447
BOD ₅ (mg/L)	1.13291(1.03447)	1.08433(1.06099)	0.324
TP (mg/L)	0.0762(0.05613)	0.06301(0.0558)	0.825
Petroleum (mg/L)	0.00448(0.00506)	0.00547(0.01616)	<.001
DO (mg/L)	8.47957(1.54288)	8.11044(1.82962)	<.001
Cyanide (mg/L)	0.00078(0.00083)	0.00074(0.00045)	<.001
Social Factors			
Per capita gross domestic product (GDP) (10 000 yuan)	6.01(2.78)	5.08(2.72)	0.003
Fraction of the population aged 65 years and older	14.62(3.21)	12.17(3.78)	<.0001
Urbanisation rate	57.95(12.79)	53.44(15.7)	0.007

Note: The p-value was calculated using the rank sum test.

Supplementary Information Code.

Supplementary Note 1. Code for conducting negative binomial regression.

```
library(MASS)
mean<-t(aggregate(data[,cancername],by=list(data$jgrade),FUN=mean))
data[, "jgrade"]<-as.factor(data[, "jgrade"])
model<-list()
model[[1]]<-glm.nb(Oesophageal~jgrade,data=data)
model[[2]]<-glm.nb(Stomach~jgrade,data=data)
model[[3]]<-glm.nb(Colorectal~jgrade,data=data)
model[[4]]<-glm.nb(Liver~jgrade,data=data)
model[[5]]<-glm.nb(Gallbladder~jgrade,data=data)
model[[6]]<-glm.nb(Pancreatic~jgrade,data=data)
model[[7]]<-glm.nb(Lung~jgrade,data=data)
model[[8]]<-glm.nb(Bone~jgrade,data=data)
model[[9]]<-glm.nb(Breast~jgrade,data=data)
model[[10]]<-glm.nb(Kidney~jgrade,data=data)
model[[11]]<-glm.nb(Brain~jgrade,data=data)
model[[12]]<-glm.nb(Leukemia~jgrade,data=data)
model[[13]]<-glm.nb(Lymphoma~jgrade,data=data)
model[[14]]<-glm.nb(Total~jgrade,data=data)
names(model)<-cancername
RR<-mean[2:15,]
RR[,1]<-1
for(c in 1:14)
  for(j in 1:3)
  {
    RR[c,j+1]<-exp(summary(model[[c]])$coefficients[j+1,1])
  }
RR_lower<-mean[2:15,]
RR_lower[,1]<-1
for(c in 1:14)
  for(j in 1:3)
  {
    RR_lower[c,j+1]<-exp(confint(model[[c]])[j+1,1])
  }
RR_upper<-mean[2:15,]
RR_upper[,1]<-1
for(c in 1:14)
  for(j in 1:3)
  {
    RR_upper[c,j+1]<-exp(confint(model[[c]])[j+1,2])
  }
```

The code runs in R and relies on the MASS package.

Supplementary Note 2. Code for producing forest map.

```
library(forestplot)
tabletext <- cbind(c("Cancer site",as.character(rs_forest$Cancersite)),
                  c("Mean" ,as.character(rs_forest$Cancerrate_mean)),
                  c("RR(95%CI)" ,as.character(rs_forest$CI)))
forestplot(labeltext = tabletext,
           mean = c(NA,rs_forest$RR),
           lower =c(NA,rs_forest$Lower),
           upper = c(NA,rs_forest$Upper),
           is.summary=c(T,T,F,F,F,T,F,F,F,T,F,F,F,T,F,F,F,T,F,F,F,T,F,F,F,T,F,F,F,T,F,F,F),
           align=c('c','c','c'),
           xticks = c(1/2,1,3),
           colgap = unit(4,"mm"),
           zero = 1,
           boxsize = 0.3,
           lineheight = unit(4,'mm'),
           graphwidth = unit(.3,"npc"),
           lwd.zero = 2,
           lwd.ci = 1,
           col=fpColors(box='#BC3C28',summary="#BC3C28",lines = 'black'),
           xlab="",
           lwd.xaxis=1,
           lty.ci = "solid",
           txt_gp = fpTxtGp(ticks=gpar(cex=0.6,fontface="bold"),
                           label = list(gpar(fontface="bold",col = "#BC3C28",cex=0.6),
                                         gpar(cex=0.6),
                                         gpar(cex=0.6),
                                         gpar(cex=0.6),
                                         gpar(cex=0.6)
                                         )
                           ),
           graph.pos = 4,
           hrzl_lines=list("2" = gpar(lwd=1, col="grey3"))
)
```

The code runs in R and relies on the forestplot package.

Supplementary Note 3. Code for Conducting shap analysis and producing shap summary plot.

```
import sklearn
import shap
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import RandomForestClassifier
for i in range (5,19):
y = dataset.iloc[:, i]
rforest = RandomForestRegressor(n_estimators=1000, max_depth=None, n_samples_split=2,
    random_state = 0)
rforest.fit(X, y)
explainer = shap.TreeExplainer(rforest)
shap_values = explainer.shap_values(X)
shap.summary_plot(shap_values1, X,plot_size=0.3, max_display = 30, show=False)
```

The code runs in python and relies on sklearn and shap packages.