Article

Spatial consistency of co-exposure to air and surface water pollution and cancer in China

Jingmei Jiang [1,2,6] ✉, Luwen Zhang [1,6], Zixing Wang[1,6], Wentao Gu [1,6], Cuihong Yang [1,6], Yubing Shen[1,6], Jing Zhao[1], Wei Han[1], Yaoda Hu[1], Fang Xue [1], Wangyue Chen [1], Xiaobo Guo [1], Hairong Li[3,4], Peng Wu[1], Yali Chen[1], Yujie Zhao[1], Jin Du [1] & Chengyu Jiang[2,5] ✉

Humans can be exposed to multiple pollutants in the air and surface water. These environments are non-static, trans-boundary and correlated, creating a complex network, and significant challenges for research on environmental hazards, especially in real-world cancer research. This article reports on a large study (377 million people in 30 provinces of China) that evaluated the combined impact of air and surface water pollution on cancer. We formulate a spatial evaluation system and a common grading scale for co-pollution measurement, and validate assumptions that air and surface water environments are spatially connected and that cancers of different types tend to cluster in areas where these environments are poorer. We observe "dose–response" relationships in both the number of affected cancer types and the cancer incidence with an increase in degree of co-pollution. We estimate that 62,847 (7.4%) new cases of cancer registered in China in 2016 were attributable to air and surface water pollution, and the majority (69.7%) of these excess cases occurred in areas with the highest level of co-pollution. The findings clearly show that the environment cannot be considered as a set of separate entities. They also support the development of policies for cooperative environmental governance and disease prevention.

Air and surface water are basics for human survival[1,2]. The rapid urbanisation and industrialisation in the past century came at the price of environmental deterioration[3], which in turn, caused multiple hazards to population health[4,5]. It is estimated that over 90% of the world population in 2021, especially those in developing countries, lived in places where the World Health Organisation (WHO) standard on fine particulate matter, with a diameter of ≤ 2.5 microns ($PM_{2.5}$), was not met[6]. This is alarming because there is sufficient evidence to show that $PM_{2.5}$ has a causative association with lung cancer[7–9]. Other air pollutants are also gaining increased attention. For instance, nitrogen dioxide (NO_2) has been linked to breast cancer[10,11], and the WHO introduced this pollutant to its monitoring database in 2022[12]. Surface water is similar to air in some ways (for example, it is non-static[13], transboundary[14], and has multiple exposure routes[15,16]), and these pose challenges for research. While previous studies have investigated the impact of specific types of water pollution at a local level[17], there

[1]Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & School of Basic Medicine, Peking Union Medical College, Beijing, China. [2]Center of Environmental and Health Sciences, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China. [3]Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. [4]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China. [5]State Key Laboratory of Medical Molecular Biology, Department of Biochemistry, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences & School of Basic Medicine Peking Union Medical College, Beijing, China. [6]These authors contributed equally: Jingmei Jiang, Luwen Zhang, Zixing Wang, Wentao Gu, Cuihong Yang, Yubing Shen. ✉e-mail: jingmeijiang@ibms.pumc.edu.cn; jiang@pumc.edu.cn

remains a limited understanding of the potential health effects of exposure to combined pollutants and their association with cancer incidence. The global number of cancer cases is expected to double in approximately half a century for multiple reasons, such as population aging[18]. From a spatial perspective, studies have observed overlapping distribution patterns of various cancer types[19,20], indicating the presence of potential common environmental causes. However, the existing evidence regarding environmental carcinogenicity has been developed in a fragmented manner[21], lacking a comprehensive evaluation system for understanding the holistic relationship between the real-world environment and cancer.

Here, we hypothesised that air and surface water environments would be spatially connected and cancers of different types would tend to cluster in areas with poor environmental conditions[22,23]. To address these hypotheses and facilitate future work, we established a spatial evaluation system that harmonises nationwide data on air, surface water, and cancer incidence in China. We also developed a graded scale of co-pollution that makes it possible to transform a complex network with multiple pollutants and multiple types of cancers to enable quantifiable evaluation of the relationships. Finally, we demonstrated the graphic consistency in air and surface water quality and cancer incidence. By reshaping cross-industry monitoring data into a minable data resource, we highlight a unique opportunity to accelerate the generation of knowledge to support the development of policies for cooperative environmental governance and disease prevention.

Results

Spatial evaluation system for environment and cancer
There is considerable public concern in China about both the environment and cancer[24,25]. Different industries in the country have separately established one of the world's largest monitoring networks on air (real-time data from the China National Environmental Monitoring Centre)[26,27], surface water (monthly release through the Environmental Quality Monitoring Network)[28], and cancer (annual report from the China Cancer Registry System)[29]. These cover all the provincial-level regions that make up the Chinese mainland (Supplementary Information Fig. 1a–c). However, they differ in their spatial-temporal scales, and air, water and cancer data demonstrate considerable spatial heterogeneity across the country. These barriers prevent academics, cross-industry workers and the government from appraising the environment–cancer relationship. As a fundamental step to overcome this challenge, we "harmonised" these industry data as a Spatial Evaluation System for Environment and Cancer (SESEC, Table 1).

To spatially integrate the three types of national industry data, we defined the prefecture-level area as the basic unit. Any unit that simultaneously contained all three components (air monitoring site, water monitoring section, and cancer registry institute [CRI]) was included and denoted an analysis unit. All of these analysis units (totalling 219) constituted the study area, covering a population of 377 million (Supplementary Information Fig. 1d and Supplementary Information Table 1). For pollutants, if a specific analysis unit contained multiple monitoring points, we calculated the average value of the pollutants from these points to represent the average pollution level of that unit. This definition makes the units independent (i.e., non-overlapping both geographically and for information about the environment and outcome) for spatial analysis, having no restriction on the distribution or number of data points for the three components within one basic analytic unit, and therefore preserves the "natural" pattern of data sources to a large extent.

Items considered in the SESEC are extensive and tailored to the current situation in China (Table 1). The environmental quality items considered for assessment included six air pollutants recommended by the WHO Global Air Quality Guidelines[30], as well as 13 surface water organic pollutant indicators. These specific pollutants were chosen

based on criteria such as testability, comprehensiveness, and pollution share rate[31]. It is worth mentioning that surface water metal pollutants were not included in the assessment because they are already subjected to strict control measures in China and are present at very low concentrations. Threshold concentrations for the environmental quality items were set using the annual limit values recommended by the National Ambient Air Quality Standard for air pollutants[26,32]. They were based on the 75th percentiles of the national levels for water pollutant indicators because there is no standard for health effect evaluation[33]. For cancer items, the SESEC included 13 cancer types, selected because of their high incidence, increasing trend or low survival rate[34,35]. Collectively, these cancers accounted for 77% of all new cases in 2016. All the items were based on their annual average levels (with a few exceptions), to smooth temporal fluctuations (e.g., seasonal effect) and therefore better reflect a relative "stable" exposure–outcome relationship[36].

We observed very high cross-country heterogeneity in the concentration levels of all 19 environmental items, the incidence of 13 types of cancer and high spatial autocorrelation among pollutants (descriptive data in Supplementary Information Tables 2, 3). There were some high correlations within and across the air, surface water and cancer elements in the system (Supplementary Information Fig. 2). For instance, the Spearman's correlation coefficients were up to 0.88 within air pollutants ($PM_{2.5}$ and particulate matter with diameter ≤ 10 microns [PM_{10}]), 0.75 within water pollutants (permanganate index [COD_Mn] and total phosphorus [TP]), and 0.50 between air and water pollutants (PM_{10} and total nitrogen [TN]). The incidence of types of cancer also had similar patterns, with correlation coefficients up to 0.72 (breast and kidney cancer). Most air and water pollutants were also correlated with cancers. For instance, the correlation coefficient was 0.44 between PM_{10} and oesophageal cancer and 0.34 between COD_Mn and lung cancer. These figures provide a preliminary glance at the complex network in the SESEC and serve as primary proof of the inter-dependency between the air and surface water environments and human cancer.

Distribution of co-pollution measured on a common graded scale
The environment as a whole is very difficult to understand, given the large heterogeneity and complex relationships that exist both within and across environmental media. This hinders progress in the exploration of the relationship between the environment and diseases such as cancer. There are separate standards for classifying the quality of air or surface water, for example, the Air Quality Index (AQI)[37] or Water Quality Index (WQI)[38], but these classification systems use single-factor evaluation (i.e., the single pollutant with the highest pollution level), and do not take into account the spatial relation between monitoring sites or sections[39]. There is also no measurement tool to quantify and compare the degree of co-pollution in different places. We, therefore, proposed an approach to translate the complex environmental network relationship into common graded scales so as to quantify their combined effects on cancer occurrence.

The pollution grade was achieved in three successive steps. (1) For each pollutant, using its threshold concentration as the cut-off value, we used a modified local Moran's index to identify the aggregation characteristics of the spatial distribution of the pollutant. (2) Assuming a high correlation among pollutants, we applied the principle of "combining items with similar features" to facilitate the assessment of the combined effect of multiple pollutants. (3) Based on the combination of similar features, the whole space was divided into four progressive grades, enabling the transformation of a multi-dimensional complex network into a one-dimensional quantifiable co-pollution grade.

Specifically, to identify spatial patterns, we used the threshold concentration assigned to each pollutant item (as shown in Table 1)

Table 1 | Detailed information about the Spatial Evaluation System for Environment and Cancer (SESEC)

	Air	Surface water	Cancer
Data source	The China National Environmental Monitoring Centre (CNEMC)	Ministry of Ecology and Environment of China	China Cancer Registry Annual Report (2019)
Items	$PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO	Volatile phenol, Sulphate, F^-, AS, TN, NH_3-N, COD_Mn, COD, BOD_5, TP, Petroleum, DO, Cyanide	Oesophageal, stomach, colorectal, liver, gallbladder, pancreatic, lung, bone, breast, kidney, and brain cancers, leukaemia, and lymphoma
Spatial scales	1436 air monitoring sites in 331 prefecture-level areas	3632 water monitoring sections covering all nine big river basins	487 cancer registry institutes (CRIs)
Time scales	Real-time data in 2015	Monthly average for the 12 months of 2021	Annual incidence in 2016
Time scales for analysis	Peak season concentration of O_3 and the annual average concentration of the remaining five air pollutants ($PM_{2.5}$, PM_{10}, NO_2, SO_2, CO)	The annual average concentration of the 13 water pollutants	Annual
The number of components contained in the study area	219 out of 331 prefecture-level areas were matched with CRIs.	Surface water monitoring sections were matched to 219 prefecture-level areas with CRIs. The matching design used each water quality monitoring section as the "centre" of a buffer zone with a 30 km radius. It retained 2053 out of 3632 water monitoring sections.	477 out of 487 CRIs
Threshold concentration	We used the annual limit values recommended by the National Ambient Air Quality Standard of China ($PM_{2.5}$: 35 µg/m³; PM_{10}: 70 µg/m³; NO_2: 40 µg/m³; SO_2: 60 µg/m³) as threshold concentration. For the air pollutants without annual limit values given (CO and O_3), the proportion of eligible cities reported in Bulletin on the State of China's Environment in 2015 was applied, i.e., the 96.7th percentile for CO and the 84th percentile for O_3 were considered to be the threshold concentration.	75th percentile of each pollution indicator value was used as a threshold for defining inferior water quality levels.	

$PM_{2.5}$, particulate matter ≤ 2.5; PM_{10}, particulate matter ≤10; NO_2, nitrogen dioxide; O_3, ozone; SO_2, sulphur dioxide; CO, carbon monoxide; F^-, fluoride; AS, anionic surfactant; TN, total nitrogen; NH_3-N, ammonia nitrogen; COD_Mn, permanganate index; COD, chemical oxygen demand; BOD_5, biochemical oxygen demand after 5 days; TP, total phosphorus; DO, dissolved oxygen.

Fig. 1 | Results of local Moran's *I* of six air pollutants and 13 surface water pollutants. $PM_{2.5}$, particulate matter ≤ 2.5; PM_{10}, particulate matter ≤10; NO_2, nitrogen dioxide; O_3, ozone; SO_2, sulphur dioxide; CO, carbon monoxide; F^-, fluoride; AS, anionic surfactant; TN, total nitrogen; NH_3-N, ammonia nitrogen; COD_Mn, permanganate index; COD, chemical oxygen demand; BOD_5, biochemical oxygen demand after 5 days; 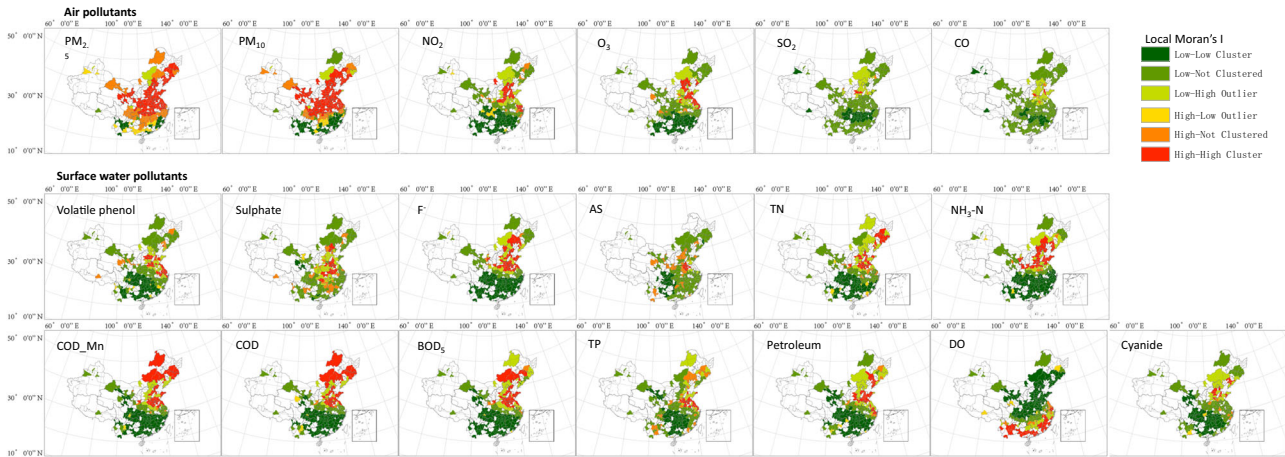TP, total phosphorus; DO, dissolved oxygen. The base map was obtained from the China Resource and Environment Science and Data Centre.

and calculated a modified local Moran's I[33,40]. Through conversion to a binary variable, the modified local Moran's index is intrinsically equivalent to converting the original local Moran's index with the mean as the cut-off value to the threshold concentration as the cut-off value, such that we can identify the high value in the sense of threshold concentration specific to each pollutant. This index helped us identify six various cluster patterns, including: a high–high cluster (HH), low–low cluster (LL), high–low outlier (HL), low–high outlier (LH), high–not clustered (HN), and low–not clustered (LN) (detailed in the Methods). This standardisation process treats all pollutants equally, regardless of their measurement units. It also considers the non-static and transboundary characteristics of air and water pollutants. The geographical details regarding clustering or outliers of pollutants (as shown in Fig. 1) can provide valuable insights for regionally tailored environmental policy-making. We focused on the high-level pollutants ("H" for short) and counted the number of "H"s to grade the degree of pollution in air or surface water. Different grades across environmental media can be further tabulated as a matrix to show patterns of co-pollution (Fig. 2a). This matrix network could also be simplified as a common
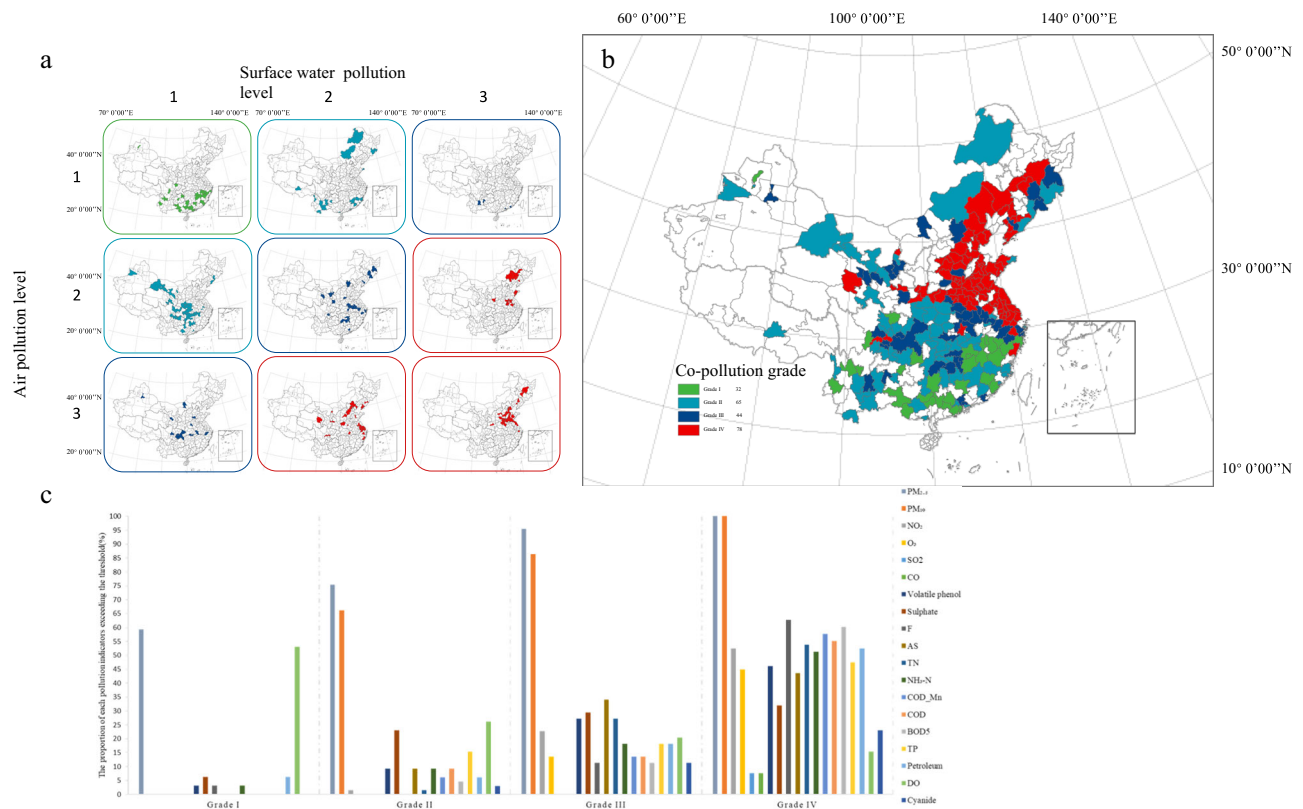
Fig. 2 | Spatial distribution of co-pollution grades and composition of pollutants at each grade. Spatial distribution of air and surface water pollution level (**a**) and their derived co-pollution grade (**b**), and the proportion of each type of pollutant exceeding the threshold concentrations (**c**). Six types of spatial patterns were defined using the results of hypothesis testing for local Moran's *I* statistic: high–high cluster (HH, indicating a unit with high pollutant levels surrounded by neighbouring units with similarly high levels, exceeding the defined threshold), low–low cluster (LL), high–low outlier (HL), low–high outlier (LH), high–not clustered (HN) and low–not clustered (LN). For any analytic unit, pollutants with a spatial distribution pattern of HH, HL, or HN were defined as high-level air pollutants ("H"). For air, units with 0–1, 2, and 3–6 "H" pollutants were defined as 1, 2, and 3 (low- to high-level pollution), respectively. For water, units with 0–1, 2–5, and 6–13

"H" pollutants were defined as 1, 2, and 3 (low- to high-level pollution). Based on the 3 × 3 crossed table of the levels of both air and surface water pollution, we defined the co-pollution grade: Grade I (both air and water pollution at low level,1-1); Grade II (1-2 or 2-1); Grade II (1-3,3-1 or 2-2); and Grade IV (2-3,3-2 or 3-3). $PM_{2.5}$, particulate matter ≤ 2.5; PM_{10}, particulate matter ≤10; NO_2, nitrogen dioxide; O_3, ozone; SO_2, sulphur dioxide; CO, carbon monoxide; F^-, fluoride; AS, anionic surfactant; TN, total nitrogen; NH_3-N, ammonia nitrogen; COD_Mn, permanganate index; COD, chemical oxygen demand; BOD_5, biochemical oxygen demand after 5 days; TP, total phosphorus; DO, dissolved oxygen.Source data are provided as a Source Data file. The base map was obtained from the China Resource and Environment Science and Data Centre.

graded scale of co-pollution by merging cells with similar patterns (Fig. 2b).

Based on our empirical grouping on the common graded scale, 78 basic analytic units (35.6%) had high-level pollution in both air and surface water, which corresponds to Grade IV on the scale. These areas were mainly distributed in the Beijing–Tianjin–Hebei region (i.e., the Capital Economic Circle), the Huaihe River basin (which has a dense water network and population) and the Fen-Wei Plain (downstream of the Yellow River). All 19 pollutants exceeded the thresholds in these areas, with the exposure rate (proportion of basic analytic units exceeding the threshold for each pollutant) of $PM_{2.5}$ and PM_{10} approaching 100% (Fig. 2c). At the other extreme, 32 (14.6%) basic analytic units had very low-level pollution of both air and surface water (Grade I). In these areas, all in southern China, only 7 pollutants exceeded the thresholds, and the exposure rate was low, with the highest exposure rate of 59.4% for $PM_{2.5}$. In between, 65 (29.7%) and 44 (20.1%) basic analytic units were classified as Grade II (low-level pollution in both) and III (moderate-level co-pollution), respectively. Note that very few areas had either high-level pollution in the air but low-level pollution in surface water (11 [5.0%] basic analytic units, scattered in Northern and Central China) or the opposite (four [1.8%] basic analytic units, in border areas in the southeast and southwest). This reinforces the spatial connection between air and surface water

pollution and the validity of the proposed grading system of co-pollution.

We quantified the degrees of pollution uniformly and simultaneously in both air and surface water, and to show a spatial connection between them. Despite that, the final grade of co-pollution was affected by the thresholds used for the pollutants (e.g., the number of areas of a higher grade would be reduced if less strict criteria, say 80th percentiles, was used for water pollutants), the overall pattern would not change, and the present results are supported by some previous knowledge of environmental problems. For example, the Grade IV areas are mainly distributed in populated regions where air pollution (such as the Beijing–Tianjin–Hebei region) or water pollution (such as the Huaihe River basin) have aroused public concerns[41,42]. The pollution may be due to gases, wastewater and solid waste from the chemical-based industrial structure, the road freight-based transport structure and the coal-based energy structure[43]. Our results further stress the co-pollution problems in these areas, i.e., the possibility of shared pollution sources for both air and surface water. These suggest that coordinated governance across sectors is required to balance economic development and the environment. China is also a miniature of the discrepancy in both pollution degree and patterns that exist worldwide. No uniform development model could fit all areas. Tailored environmental policies are therefore needed.
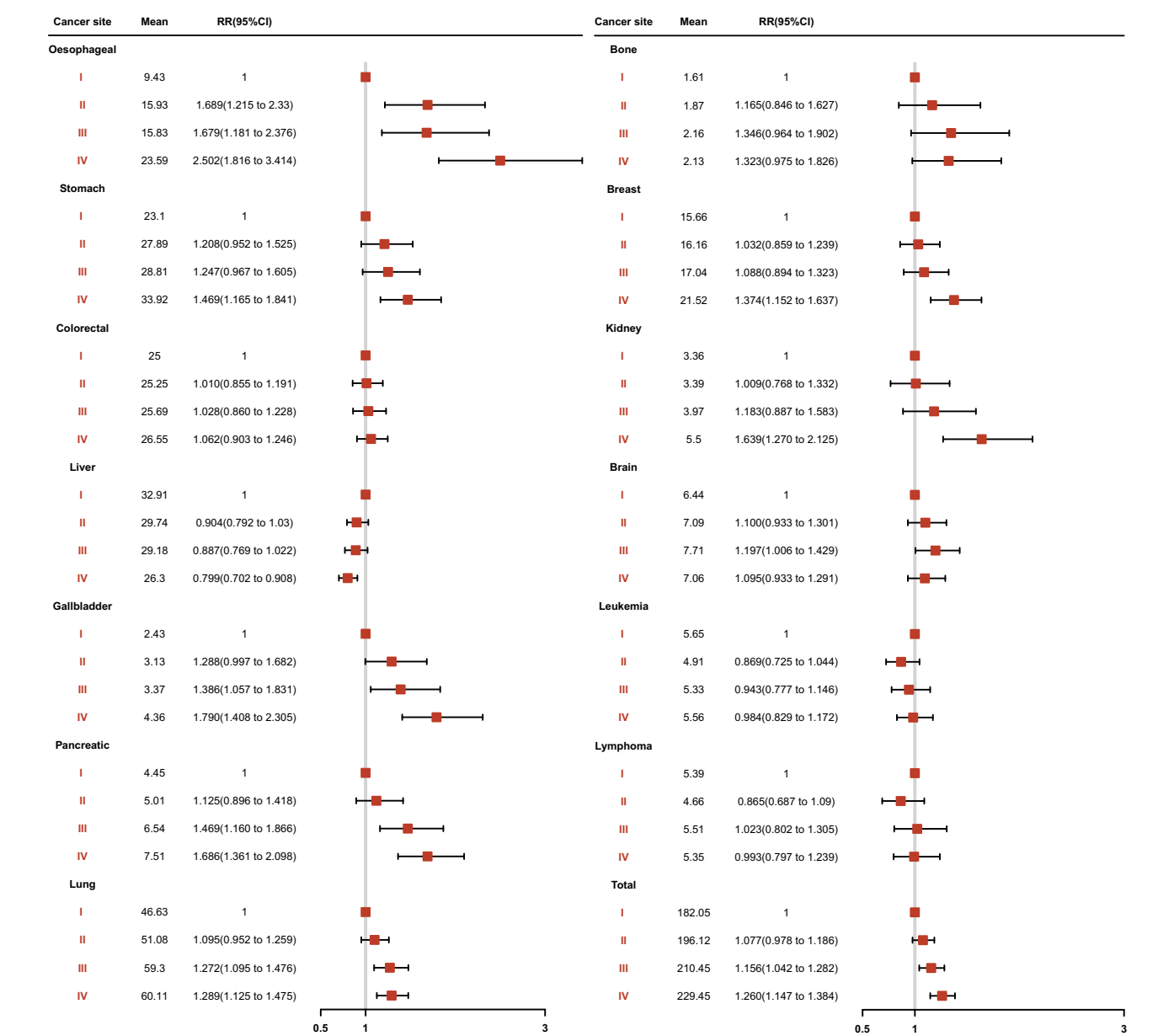
Cancer site	Mean	RR(95%CI)
Oesophageal		
I	9.43	1
II	15.93	1.689(1.215 to 2.33)
III	15.83	1.679(1.181 to 2.376)
IV	23.59	2.502(1.816 to 3.414)
Stomach		
I	23.1	1
II	27.89	1.208(0.952 to 1.525)
III	28.81	1.247(0.967 to 1.605)
IV	33.92	1.469(1.165 to 1.841)
Colorectal		
I	25	1
II	25.25	1.010(0.855 to 1.191)
III	25.69	1.028(0.860 to 1.228)
IV	26.55	1.062(0.903 to 1.246)
Liver		
I	32.91	1
II	29.74	0.904(0.792 to 1.03)
III	29.18	0.887(0.769 to 1.022)
IV	26.3	0.799(0.702 to 0.908)
Gallbladder		
I	2.43	1
II	3.13	1.288(0.997 to 1.682)
III	3.37	1.386(1.057 to 1.831)
IV	4.36	1.790(1.408 to 2.305)
Pancreatic		
I	4.45	1
II	5.01	1.125(0.896 to 1.418)
III	6.54	1.469(1.160 to 1.866)
IV	7.51	1.686(1.361 to 2.098)
Lung		
I	46.63	1
II	51.08	1.095(0.952 to 1.259)
III	59.3	1.272(1.095 to 1.476)
IV	60.11	1.289(1.125 to 1.475)

Cancer site	Mean	RR(95%CI)
Bone		
I	1.61	1
II	1.87	1.165(0.846 to 1.627)
III	2.16	1.346(0.964 to 1.902)
IV	2.13	1.323(0.975 to 1.826)
Breast		
I	15.66	1
II	16.16	1.032(0.859 to 1.239)
III	17.04	1.088(0.894 to 1.323)
IV	21.52	1.374(1.152 to 1.637)
Kidney		
I	3.36	1
II	3.39	1.009(0.768 to 1.332)
III	3.97	1.183(0.887 to 1.583)
IV	5.5	1.639(1.270 to 2.125)
Brain		
I	6.44	1
II	7.09	1.100(0.933 to 1.301)
III	7.71	1.197(1.006 to 1.429)
IV	7.06	1.095(0.933 to 1.291)
Leukemia		
I	5.65	1
II	4.91	0.869(0.725 to 1.044)
III	5.33	0.943(0.777 to 1.146)
IV	5.56	0.984(0.829 to 1.172)
Lymphoma		
I	5.39	1
II	4.66	0.865(0.687 to 1.09)
III	5.51	1.023(0.802 to 1.305)
IV	5.35	0.993(0.797 to 1.239)
Total		
I	182.05	1
II	196.12	1.077(0.978 to 1.186)
III	210.45	1.156(1.042 to 1.282)
IV	229.45	1.260(1.147 to 1.384)

Fig. 3 | Estimated rate ratios and 95% confidence intervals for incidence of 13 types of cancer in relation to co-pollution grades. I, Grade I ($N = 32$); II, Grade II ($N = 65$); III, Grade III ($N = 44$); IV, Grade IV ($N = 78$). Source data are provided as a Source Data file.

Cancer incidence in relation to environmental pollution

Stephen Paget studied the patterns of cancer metastasis and then proposed the Seed-and-Soil Theory[44]. This states that metastasis depends on interactions between cancer cells (the 'seeds') and specific organ microenvironments (the 'soil') and that cancer cells exhibit preferences when metastasising to organs. We assumed that cancers, when viewed from the spatial perspective, also have preferences for particular environmental conditions, i.e., they tend to cluster in areas with particular environmental characteristics. We examined whether cancers of different types display similar spatial patterns in the population.

Interestingly, we found good consistency in the spatial distributions between the cancer incidence and the co-pollution grade. The spatial consistency was especially clear for lung, stomach and oesophageal cancers, the three most common cancers in China (Supplementary Information Fig. 3)[34].

To provide some insights about the spatial consistency, we showed that Grade IV areas had the highest levels of incidence of seven types of cancer, including oesophageal (incidence rate ratio [RR] of 2.502 compared to Grade I, an increase in risk of 150.2%), gallbladder (1.790), pancreatic (1.686), kidney (1.639), stomach (1.469), breast (1.374), and lung (1.289). In Grade II and III areas, the incidence of one and five types of cancer, respectively, was significantly higher than in Grade I areas (Fig. 3). There was a "dose-response" relationship between the number of affected cancer types and the cancer incidence with an increase in co-pollution grade. This relationship remained consistent across different grouping schemes used to define the co-pollution grade, as indicated in Supplementary Information Fig. 4. This sensitivity analysis further strengthens the evidence supporting the combined impact of environmental conditions on cancer outcomes.

Looking at the effect of specific pollutants on specific cancers, all 19 pollutants had potentially important effects on at least one cancer type filtered by Shapley additive explanations (SHAP) analysis[45] (Fig. 4). Among these, eight pollutants (four air pollutants, PM_{10}, $PM_{2.5}$, NO_2 and ozone [O_3], and four water pollutants, COD_Mn, petroleum, dissolved oxygen [DO], cyanide) showed significant positive effects
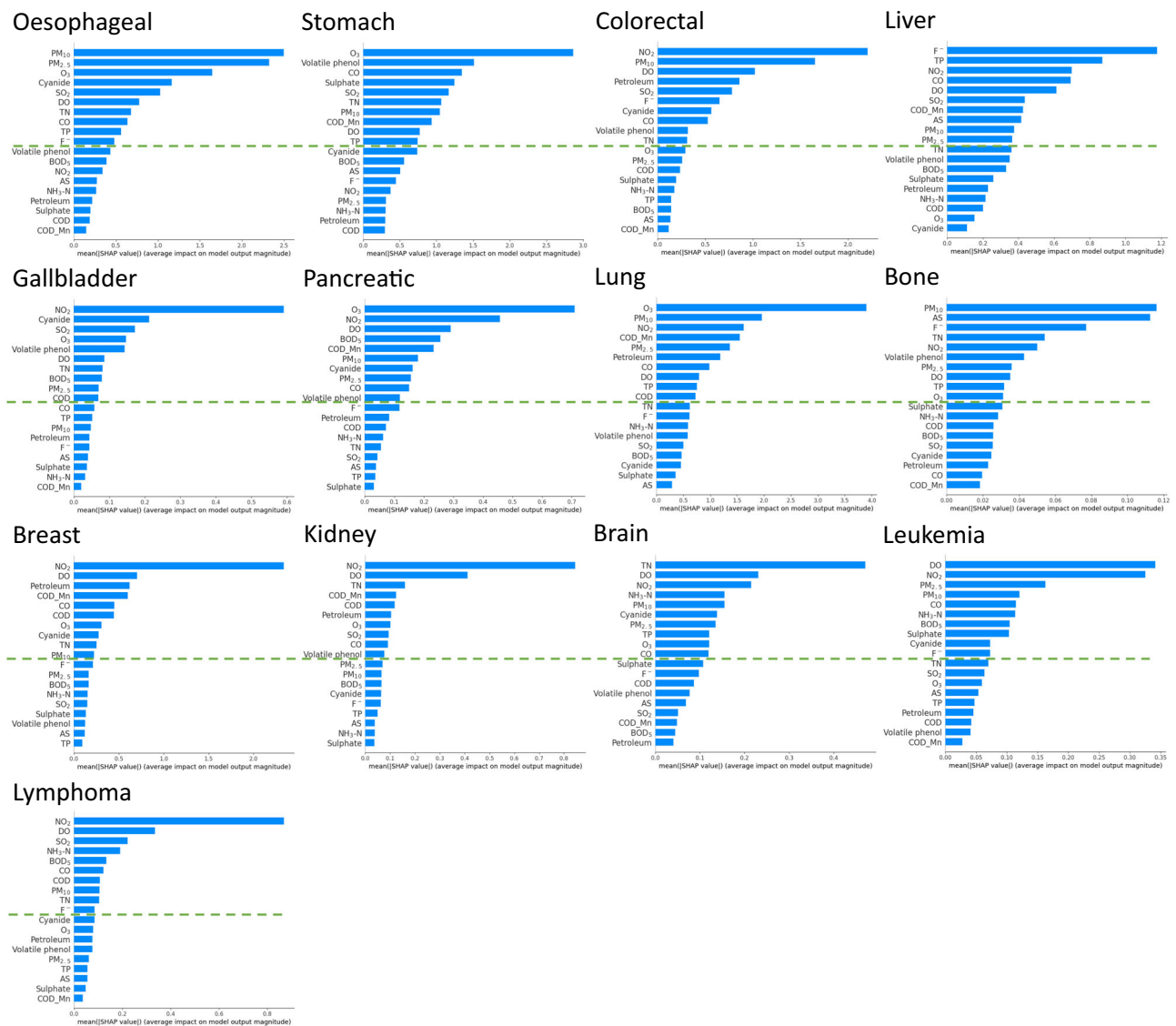
Fig. 4 | Contribution ranking of the pollutants to each cancer type, evaluated by SHAP values. The red line marks the top 10 pollutants regarding their SHAP values for the cancer type. PM$_{2.5}$, particulate matter ≤ 2.5; PM$_{10}$, particulate matter ≤ 10; NO$_2$, nitrogen dioxide; O$_3$, ozone; SO$_2$, sulphur dioxide; CO, carbon monoxide;

F$^-$, fluoride; AS, anionic surfactant; TN, total nitrogen; NH$_3$-N, ammonia nitrogen; COD_Mn, permanganate index; COD, chemical oxygen demand; BOD$_5$, biochemical oxygen demand after 5 days; TP, total phosphorus; DO, dissolved oxygen. Source data are provided as a Source Data file.

(Table 2). The per capita gross domestic product, the fraction of the population aged 65 years and older, and the urbanisation rate were also identified as significant contributors to cancer risks (spatial patterns presented Supplementary Information Fig. 5). After adjusting for these social factors, the observed effects of the pollutants remained stable (Supplementary Information Fig. 6). However, the study found that there was no positive correlation between natural environments and liver cancer after adjusting for these social factors. This suggests that social factors may be more important than natural environments in terms of liver cancer, which is primarily driven by hepatitis B and C infections in China[46].

A relationship between NO$_2$ and breast cancer has been established[47], and some previous studies pointed to a similar relationship with colorectal cancer[47] and leukaemia[48]. We extended these findings to nine cancer types, including colorectal (RR = 1.132), gallbladder (1.102), pancreatic (1.172), lung (1.042), breast (1.119), kidney (1.126), and brain (1.056) cancers, leukaemia (1.099) and lymphoma (1.233). These observations could be used to reinforce the rationality for including NO$_2$ in the WHO ambient air quality database[12]. We

confirmed that PM$_{2.5}$ has a causal relationship with lung cancer (RR = 1.188). Our findings also suggest this known Type I carcinogen[21] may have an effect on leukaemia (RR = 1.298). We also observed a relation between COD_Mn and three types of cancers, including pancreatic (RR = 1.089), breast (1.274), and kidney (1.177). COD_Mn is extensively utilised in China as a comprehensive indicator for assessing nitrite and organic pollutants in surface water[49]. The combination of nitrite with amines can generate nitrosamine, which is a known carcinogen. In addition, direct carcinogens and pre-carcinogens are organic substances that have the potential to induce DNA changes[50,51]. This biological basis supports our findings. In situations where testing capacity is insufficient, our study suggests that COD_Mn could serve as a simplified indicator for assessing both nitrite and organic pollutants, thereby providing valuable information on cancer risks associated with water pollution.

Recognising that pollutants do not exist in isolation, it is important to acknowledge that these substances also do not act independently. Even pollutants that have not yet been acknowledged as carcinogenic can still have an impact on the risk of cancer in

Table 2 | The estimated rate ratios and 95% confidence intervals for incidence effect of cancer-specific pollutants on each cancer

		Oesophageal	Colorectal	Gallbladder	Pancreatic	Lung	Bone	Breast	Kidney	Brain	Leukaemia	Lymphoma
PM$_{2.5}$	M1					1.22 [1.11, 1.35]					1.33 [1.13, 1.56]	1.38 [1.27, 1.49]
	M2					1.19 [1.09, 1.30]					1.30 [1.10, 1.54]	1.23 [1.12, 1.36]
PM$_{10}$	M1	1.63 [1.18, 2.27]					1.43 [1.09, 1.87]					
	M2	1.86 [1.37, 2.53]					1.39 [1.05, 1.84]					
NO$_2$	M1		1.24 [1.17, 1.32]	1.25 [1.13, 1.38]	1.30 [1.20, 1.41]	1.11 [1.05, 1.17]		1.19 [1.11, 1.27]	1.25 [1.13, 1.38]	1.11 [1.03, 1.19]	1.14 [1.05, 1.23]	
	M2		1.13 [1.07, 1.20]	1.10 [0.98, 1.24]	1.17 [1.07, 1.28]	1.04 [0.99, 1.10]		1.12 [1.04, 1.21]	1.13 [1.02, 1.24]	1.06 [0.97, 1.15]	1.10 [1.00, 1.21]	
O$_3$	M1			1.10 [1.01, 1.20]	1.16 [1.08, 1.25]	1.08 [1.03, 1.13]						
	M2			1.05 [0.96, 1.15]	1.07 [1.00, 1.15]	1.04 [1.00, 1.08]						
COD_Mn	M1				1.15 [1.05, 1.25]			1.32 [1.14, 1.54]	1.30 [1.04, 1.61]			
	M2				1.09 [1.00, 1.18]			1.27 [1.10, 1.47]	1.27 [1.10, 1.47]			
Petroleum	M1		1.07 [1.02, 1.13]			1.05 [1.01, 1.10]		1.09 [1.03, 1.16]	1.11 [1.01, 1.21]			
	M2		1.05 [1.00, 1.10]			1.05 [1.01, 1.08]		1.09 [1.03, 1.16]	1.08 [0.99, 1.17]			
DO	M1									1.07 [1.01, 1.14]	1.09 [1.02, 1.17]	1.11 [1.03, 1.19]
	M2									1.06 [0.99, 1.13]	1.09 [1.01, 1.18]	1.13 [1.05, 1.23]
Cyanide	M1	1.13 [1.01, 1.30]		1.08 [1.01, 1.14]								
	M2	1.11 [1.00, 1.26]		1.08 [1.01, 1.14]								

M1: adjusted for other pollutants.
M2: adjusted for other pollutants and social factors.
Social factors included per capita gross domestic product, the fraction of the population aged 65 years and older, and the urbanisation rate.
PM$_{2.5}$, particulate matter ≤2.5; PM$_{10}$, particulate matter ≤10; NO$_2$, nitrogen dioxide; O$_3$, ozone; COD_Mn, permanganate index; DO, dissolved oxygen.

populations. This impact may arise from intricate interactions with known carcinogens. Findings from several previous studies have provided support for this assumption[33,52], underscoring the urgent need for further investigation into the potential network mechanisms that connect multiple pollutants and the development of cancer. Exploring these complex interactions is crucial for effectively managing the risks associated with pollutants and preventing cancer.

Acting upon the environment-attributable cancer burden
Understanding the environmental–cancer relationship is a premise for motivating actions, but this knowledge alone does not provide a sufficient basis for the formulation of environmental governance and disease prevention policies. In this section of the report, we provide estimates of the number of excess cancer cases related to air and surface water environments in areas of different co-pollution grades. This is used as a call for growth in both academic and political interest in environmental health and collaborative efforts across sectors.

Overall, there were 62,847 excess cases in the basic analytic units in 2016, which means 7.4% of total cancer cases were attributable to air and surface water pollution. As the co-pollution grade increased, the number of pollutants that could explain the excess cases increased, from three in Grade I areas to eight in Grade IV areas. The number of types of cancers that were attributable also increased, from five in Grade I areas to 10 in Grade IV areas (Fig. 5a and Supplementary Information Table 4).

The cancer spectrum attributable to different co-pollution grades was affected by the patterns of pollution (Fig. 5b). For example, PM$_{2.5}$ as a single dominating pollutant (59.4% of the basic analytic units) in Grade I areas explained 523 (4.0%) excess cases of lung cancer. In Grade II areas, 66.2% of the basic analytic units were exposed to PM$_{10}$, and there were more excess cases (1763, 11.1%) of oesophageal cancer than in Grade I areas (91, 0.58%). The number of basic analytic units exposed to NO$_2$ increased from 0.0% in Grade I and 1.5% in Grade II to 22.7% in Grade III. The excess cases of colorectal cancer and breast cancer also increased significantly, reaching 2218 (15.9%) and 1863 (13.3%) in Grade III, respectively. The Grade IV areas were exposed to the largest number of pollutants and had the highest excess cases across all cancer types, 43,827 in total (accounting for 69.7% of total excess cases). These findings could be used by local governments to scale up countermeasures.

Discussion
In this study, we integrated nationwide data on air, surface water and cancer and consolidated the methodological basic for examining the relationships between multiple pollutants and multiple types of cancers within this giant system. Data availability is fundamental to speed up future research and actions. The development of the SESEC in this study benefited from the establishment of monitoring sites, sections and CRIs across the country. Maintaining and improving coverage of this infrastructure requires considerable and costly effort. Our work provides a way to make use of the available data, and we encourage researchers in China and elsewhere to build upon knowledge that could inform environmental protection and cancer control policies. To facilitate the replication and modification of our work, we have included an overview of our study design, analytic methods, and underlying assumptions in Fig. 6. This information serves as a useful guide for researchers seeking to build upon our work and contribute to this important field.

An important aspect of the work is its contributions to the development of spatial analysis, an interdisciplinary field of geography, epidemiology, statistics and ecology. This represents a research paradigm using an ecosystem perspective, which is more macroscopic than the traditional lenses for observation, such as the molecular, cellular and individual patient levels. Because our 219
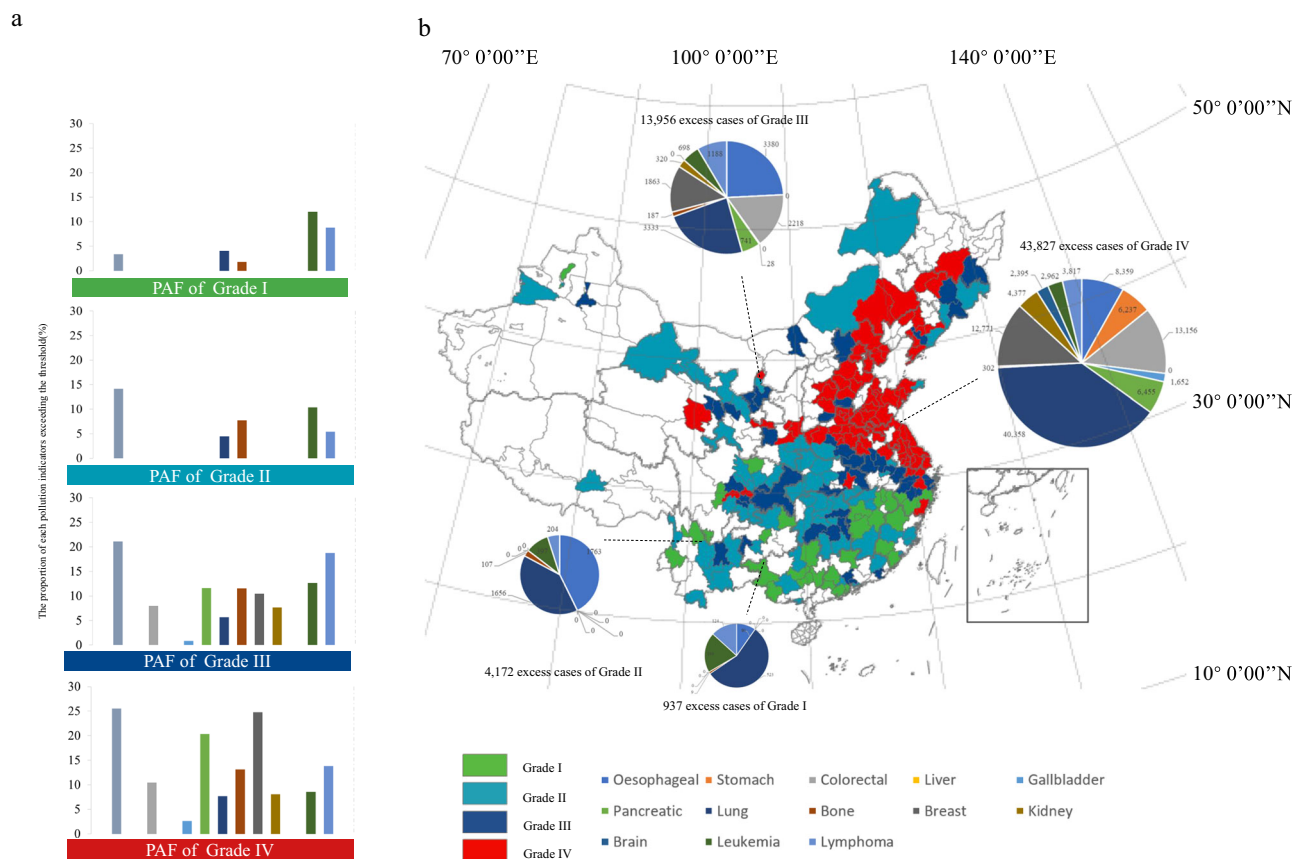
Fig. 5 | Cancer burden attributable to pollution. Proportion of cancer cases attributable to overall pollutants in each grade (**a**) and excess cases of different cancers in each co-pollution grade area (**b**). The vertical axis in (**a**) shows the proportion of cases of different types of cancer attributable to pollutants in each co-pollution grade area. PAF: population attributable fraction. Source data are provided as a Source Data file. The base map was obtained from the China Resource and Environment Science and Data Centre.
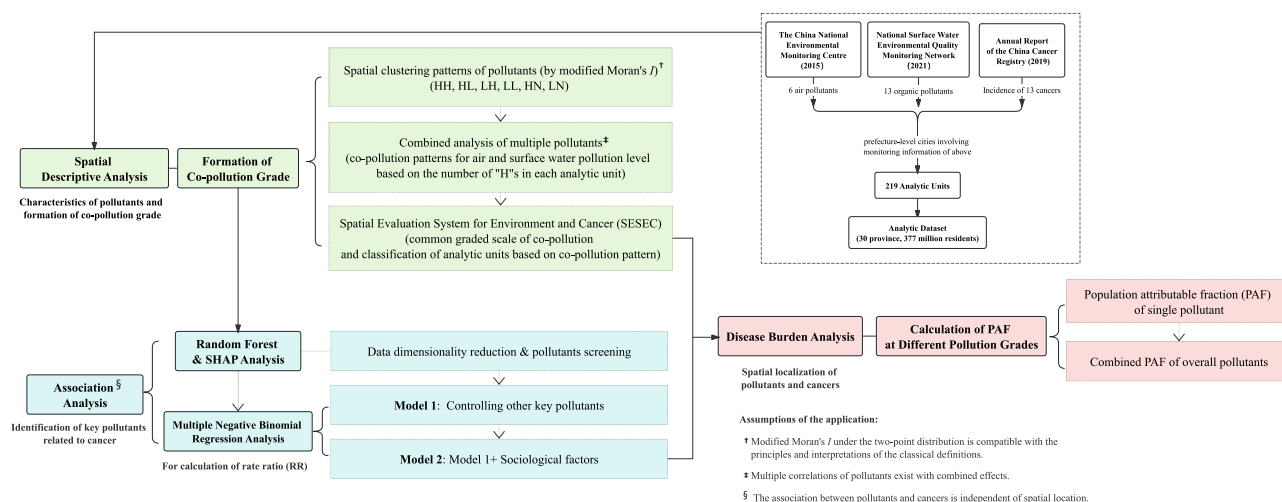
Fig. 6 | An overview of the study design, analytic methods, and underlying assumptions. This figure clarifies each part of the work and their logical connections.

analytical units covered nearly all CRIs (477/487), our evaluation of cancer risk is unbiased. As anticipated, however, for all non-included units, the levels of air, water, and related sociological factors were significantly lower than those in the study area (Supplementary Information Table 5). This phenomenon illustrates the focus of national monitoring data and also highlights that coordinating economic development and environmental governance in less-developed areas will be a more effective choice.

The unique nature of air and surface water as environments to which individuals are continuously exposed from birth presents several challenges when attempting to establish associations between these exposures and cancer outcomes. Challenges include factors such as the lack of quantification approaches for individual exposure levels, the intricate interplay between genetic and environmental factors and unknown time lags. In addressing these challenges, spatial analysis offers an alternative perspective to understand the environmental

effects. In this analysis, we emphasised the spatial consistency of data rather than the temporal sequence of the dataset. The large-scale urbanisation process and rapid economic development in China began in the early 1990s, leading to concentrated air and water pollution, particularly in the eastern coastal areas, which resulted in substantial health effects. After more than 20 years of continuous efforts, the overall air and water quality in China has been gradually improving[33]. Therefore, from a logical standpoint, correlating the occurrence of cancers several years ago with the latest water and air pollution data would likely underestimate the risks. We have reason to believe that the connection effect between the poorer water and air quality in China 10 or 15 years ago and the subsequent occurrence of cancers would be stronger. Furthermore, it is important to note that no country's administrative data are specifically designed for a particular research topic. However, certain studies (such as ours) must rely on national-level data to be more credible and comprehensive. This is because it is difficult to encompass the health effects of air and water on even large populations through population-based studies.

This study encompasses a vast majority of China's geographical area, with a population of 377 million. As a result, the data on the macroscopic system can be viewed as "parameters" reflecting the environmental conditions and population outcomes in specific locations. These parameters exhibit relative stability over time, as demonstrated in Supplementary Information Figs. 7, 8. By utilising these parameters, we can bypass the limitations inherent in population-based studies and gain valuable insights into the relationships between environmental factors and cancer outcomes. The wide range in geography that far exceeds human impact also means that these parameters demonstrated significant variations (i.e., randomness) that can help reveal interesting patterns that may suggest causation. However, establishing causation for cancer is very challenging, and will require close cooperation between different industries and disciplines to effectively control the increasing disease burden.

Methods

Data source and processing

We obtained data on air pollutants, surface water contaminants and population cancer incidence from the CNEMC, Ministry of Ecology and Environment, and China Cancer Registry Annual Report (2019), respectively.

The establishment of the air quality monitoring stations followed the guidelines outlined in the Technical Regulation for Selection of Ambient Air Quality Monitoring Stations[53]. The selection of monitoring stations follows the principles of representativeness, comparability, integrity, foresight and stability, taking into account factors such as environmental and socio-economic characteristics comprehensively (information about each monitoring station is available elsewhere[27]).

The establishment of the surface water monitoring sections followed the guidelines outlined in the National Technical Specifications for Surface Water Environmental Monitoring[54]. These sections were strategically chosen to accurately represent the natural characteristics of river network density, runoff supply, and hydrological features. For records that fell below the limits of detection (LOD), we adopted 1/2 LOD for processing according to the technical specification for surface water quality assessment published by the Ministry of Ecology and Environment of the People's Republic of China[55]. In some monitoring sections (ranging from 1.8% to 9.2% per month), data were unavailable owing to factors such as dry conditions, freezing, or other reasons. The proportion of missing data was small and it presented spatial randomness, so it may have little influence on the overall effect estimation. To address this issue and ensure a comprehensive analysis, we used averaging techniques over the analytic year for each specific section. By calculating the average concentrations over the course of the year, we also aimed to smooth out temporal fluctuations, which are

common in air and surface water pollutants, and provide a more comprehensive reflection of the potential long-term effects on cancer incidence. To establish high-value status and determine thresholds for multiple water quality pollutants, we employed a practical criterion of computing the 75th percentile values for each pollutant. This approach provides sufficient variability in various pollutant indicators and their binary conditions when performing high-value analysis. Furthermore, using the 75th percentile as a threshold offers a uniform way to determine thresholds for multiple pollutants, as we have observed their correlation with multiple types of cancer[33].

To account for social determinants of cancer risk, we obtained data on several key factors from China Statistical Yearbook 2020[56]. Specifically, we looked at per capita gross domestic product (GDP), the fraction of the population aged 65 years and older, and the urbanisation rate. The base map was obtained from the China Resource and Environment Science and Data Centre[57].

Modified local Moran's I index for identification of spatial clustering patterns

We utilised a modified version of Moran's I, originally designed for continuous variables, to analyse the spatial clustering patterns of high-value status regarding the pollutants[33]. This modified Moran's I account for categorical variables. The formula used to calculate this modified Moran's I is as follows:

$$I_i = \frac{x_i^* - \bar{x}^*}{S_i^2} \times \sum_{j=1, j\neq i}^{n} w_{ij}(x_j^* - \bar{x}^*), \quad S_i^2 = \frac{\sum_{j=1, j\neq i}^{n}(x_j^* - \bar{x}^*)^2}{n-1}, \quad (1)$$

where x^* is a binary variable that takes on values of 0 or 1. It serves as an indicator variable to determine whether the concentration x exceeds a certain threshold. If the concentration x exceeds the threshold, $x^* = 1$; otherwise, if the concentration x does not exceed the threshold, $x^* = 0$. Thus it follows a two-point distribution. $\bar{x}^*$ is the mean of the binary variable. n is the number of analytic units. x_i^* and x_j^* represent the values of the binary variable for the ith and jth analytic units, respectively. The weight w_{ij} is defined as the inverse of the distance between neighbouring units i and j. To calculate this weight, we used the minimum distance that ensures each unit has at least one neighbouring unit as the distance threshold.

The same as the original local Moran's index, six types of spatial patterns were derived based on the value of the local variable, the value of the local Moran's index, and the results of its hypothesis testing (Table 3).

For each analytic unit, any pollutant identified as HH, HL or HN was defined as "H". One to three levels of air or surface water for each analytic unit were defined based on the number of "H"s: for air, units with 0–1, 2, and 3–6 "H" pollutants were defined as 1, 2, and 3 (low- to high-level pollution). For water, units with 0–1, 2–5, and 6–13 "H" pollutants were defined as 1, 2, and 3 (low- to high-level pollution). Thus, based on the 3 × 3 cross table of levels of both air and surface water pollution, we defined the co-pollution grade as follows: Grade I (both air and water pollution at low levels, 1-1); Grade II (1-2 or 2-1); Grade II (1-3, 3-1, or 2-2); and Grade IV (2-3, 3-2, or 3-3). Co-pollution was based on three criteria: order consistency, sufficient interval, and appropriate size for each group. We examined an alternative grouping scheme that met these conditions to see robustness in the dose-response relation between the co-pollution degree and cancer risk.

A mixed modelling strategy for identification of cancer-specific key pollutants

We adopted a mixed strategy of machine learning (SHAP analysis[45]) and classical statistics (negative binomial regression[58]) to identify key pollutants with an active role in the mixtures of pollutants affecting cancer incidence.

Table 3 | Definition of six clustering patterns derived from the local Moran's index

$x_i^* - \overline{x}^*$	Moran's I_i	P	Clusters	Illustration
> 0	> 0	< 0.05	High–high cluster (HH)	Neighbouring analytic units with similar high concentrations, clustered together
> 0	< 0	< 0.05	High–low outlier (HL)	Analytic units with a high concentration surrounded by others with low concentrations
> 0		≥ 0.05	High–not clustered (HN)	High-concentration analytic units but no significant clusters
< 0	> 0	< 0.05	Low–low cluster (LL)	Opposite to HH
< 0	< 0	< 0.05	Low–high outlier (LH)	Opposite to HL
< 0		≥ 0.05	Low–not clustered (LN)	Opposite to HN

SHAP is a game theory-based framework that has been used to explain various supervised learning models without the need to know the exact structure inside the model. By providing both group and individual interpretations as well as information about the direction of the variable's effects on outcomes, SHAP has been widely used in medical research as a more flexible approach to model interpretation. In this study, SHAP allowed us to estimate the degree of impact of each pollutant on cancer incidence. The SHAP value ϕ_j of pollutant j was calculated as follows:

$$\phi_j = \sum_{S \subseteq F \backslash \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S)] \quad (2)$$

where F is the set of all pollutants, S is any subset of F, "| |" denotes the number of elements in the set and "!" denotes factorial. The SHAP value reflects the importance of pollutant j by calculating the weighted average of the difference between the predicted value with and without the pollutant j across all subsets S. SHAP can be based on any machine learning model, and we used the random forest algorithm. For the hyperparameters in a random forest, to fully train the model, we set the number of trees to 1000, the minimum split sample size to 2, and no restriction on the depth of the tree.

The larger the SHAP value, the greater the effect of the pollutant on the cancer incidence. Because the variability of SHAP values for variables beyond rank 10 tended to stabilise in most cancer types, to ensure an adequate number of variables and consistent screening criteria, we considered the 10 leading pollutants (approximately representing half of the total number) for each specific cancer type. This selection process enabled us to focus on the most influential pollutants in relation to cancer risk. For the selected pollutants, which were found to positively contribute to cancer risk, we conducted a multivariable negative binomial regression analysis. This regression model allowed us to quantify the magnitude of their effects by calculating the incidence rate ratio (RR). The RRs were adjusted for the presence of other pollutants, considering potential confounding effects. Furthermore, we explored the impact of social factors on cancer risk by incorporating adjustments for per capita gross domestic product, the fraction of the population aged 65 years and older, and the urbanisation rate. These adjustments were made both with and without considering the additional influence of social factors. By applying these approaches, we aimed to enhance the validity and reliability of our findings, particularly given the lack of prior knowledge regarding the specific cancer effects of the multiple pollutants under investigation.

Population attributable fraction for quantifying cancer burden
We utilised the population attributable fraction (PAF)[59] to assess the cancer attribution in each grade. This was done by using the burden attributable to the continuous exposure method to calculate the PAF for each grade separately. The PAF was calculated for each cancer type,

with the contribution of the ith pollutant in the jth (j = I–IV) grade determined using the following formula:

$$PAF_{ij} = \frac{\int_{min}^{max} RR_i(x)P_{ij}(x)dx - \int_{min}^{max} RR_i(x)P_i^*(x)dx}{\int_{min}^{max} RR_i(x)P_{ij}(x)dx} \quad (3)$$

where x represents the pollutant concentration, min and max are the minimum and maximum concentration values of each pollutant. $RR_i(x)$ is the incidence rate ratio at concentration x, which is obtained by negative binomial regression while adjusting for pollutants and social factors. The upper limit of $RR_i(x)$ is restricted to the value of the rate ratio at the threshold concentration to obtain the minimum estimate of PAF. $P_{ij}(x)$ is the concentration distribution of the ith pollutant in the jth grade. $P_i^*(x)$ is the theoretical minimum risk exposure distribution. It assumes that all units are exposed to concentrations below the threshold. By comparing the actual and assumed (referenced) conditions, we can estimate the burden.

The combined PAF of co-pollution in air and water in the jth grade was calculated as ref. 60

$$CPAF_j = 1 - \prod_i (1 - PAF_{ij}) \quad (4)$$

where PAF_{ij} is the PAF for the single pollutant calculated in (3).

ArcGIS software version 10.8 (Esri, Redlands, CA, USA) was used for the calculation of the modified local Moran's index, and to visualise all maps. Other analyses used SAS software (version 9.4), R (version 4.2.1) and package "SHAP" in Python (version 3.10). The code for the key steps can be obtained in the Supplementary Information Code (Supplementary Notes 1–3).

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Code availability

The code used in this study can be obtained in the Supplementary Information.

References

1. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. & Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **525**, 367–371 (2015).
2. Rodell, M. et al. Emerging trends in global freshwater availability. *Nature* **557**, 651–659 (2018).
3. Seifollahi-Aghmiuni, S., Kalantari, Z., Egidi, G., Gaburova, L. & Salvati, L. Urbanisation-driven land degradation and socioeconomic challenges in peri-urban areas: Insights from Southern Europe. *Ambio* **51**, 1446–1458 (2022).
4. Li, X. et al. Urbanization and health in China, thinking at the national, local and individual levels. *Environ. Health* **15**, 32 (2016).
5. Nweke, O. C. & Sanders, W. H. 3rd Modern environmental health hazards: a public health issue of increasing significance in Africa. *Environ. Health Perspect.* **117**, 863–870 (2009).
6. World Health Organization. Ambient (outdoor) air pollution. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (2022).
7. Hamra, G. B. et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environ. Health Perspect.* **122**, 906–911 (2014).
8. Hill, W. et al. Lung adenocarcinoma promotion by air pollutants. *Nature* **616**, 159–167 (2023).
9. Kulhánová, I. et al. The fraction of lung cancer incidence attributable to fine particulate air pollution in France: Impact of spatial resolution of air pollution models. *Environ. Int.* **121**, 1079–1086 (2018).
10. Amadou, A. et al. Long-term exposure to nitrogen dioxide air pollution and breast cancer risk: A nested case-control within the French E3N cohort study. *Environ. Pollut.* **317**, 120719 (2023).
11. Hystad, P. et al. Exposure to traffic-related air pollution and the risk of developing breast cancer among women in eight Canadian provinces: a case-control study. *Environ. Int.* **74**, 240–248 (2015).
12. World Health Organization. Ambient (outdoor) air pollution. Geneva, Switzerland (2016).
13. Zhang, H., et al. Changes in China's river water quality since 1980: management implications from sustainable development. *Npj Clean Water* **6**, 45 (2023).
14. Jin, H., Chen, X., Zhong, R. & Liu, M. Influence and prediction of PM2.5 through multiple environmental variables in China. *Sci. Total Environ.* **849**, 157910 (2022).
15. Lu, Y. et al. Impacts of soil and water pollution on food safety and health risks in China. *Environ. Int* **77**, 5–15 (2015).
16. Schraufnagel, D. E. et al. Air pollution and noncommunicable diseases: A review by the forum of international respiratory societies' environmental committee, part 2: Air pollution and organ systems. *Chest* **155**, 417–426 (2019).
17. Zhang, H., Li, H., Gao, D. & Yu, H. Source identification of surface water pollution using multivariate statistics combined with physicochemical and socioeconomic parameters. *Sci. Total Environ.* **806**, 151274 (2022).
18. Soerjomataram, I. & Bray, F. Planning for tomorrow: global cancer incidence and the role of prevention 2020-2070. *Nat. Rev. Clin. Oncol.* **18**, 663–672 (2021).
19. Cassetti, T., La Rosa, F., Rossi, L., D'Alò, D. & Stracci, F. Cancer incidence in men: a cluster analysis of spatial patterns. *BMC Cancer* **8**, 344 (2008).
20. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
21. WHO IARC. Agents Classified by the IARC Monographs, Volumes 1–132: WHO International Agency for Research on Cancer. (2022).
22. Ma, T. et al. China's improving inland surface water quality since 2003. *Sci. Adv.* **6**, eaau3798 (2020).
23. Han, L., Zhou, W., Pickett, S. T., Li, W. & Qian, Y. Multicontaminant air pollution in Chinese cities. *Bull. World Health Organ.* **96**, 233–242e (2018).
24. Cao, M. et al. Cancer screening in China: The current status, challenges, and suggestions. *Cancer Lett.* **506**, 120–127 (2021).
25. Liu, J. & Diamond, J. China's environment in a globalizing world. *Nature* **435**, 1179–1186 (2005).
26. Ministry of Ecology and Environment of China. Report on the State of the Ecology and Environment in China. (2015).
27. China Environmental Monitoring Station. National urban air quality real-time release platform. https://air.cnemc.cn:18007/ (2023).
28. China National Environmental Monitoring Centre. National Surface Water Environmental Quality Monitoring Network. http://www.cnemc.cn/ (2021).
29. National Cancer Center of China. Annual Report of the China Cancer Registry 2019. Beijing: China People's Medical Publishing House. (2019).
30. World Health Organization. WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. Geneva (2021).
31. Ministry of Ecology and Environment of China. Environmental Quality Standards for Surface Water. GB 3838–2002 (2002).
32. Ministry of Ecology and Environment of China. Ambient air quality standards. GB 3095–2012 (2012).
33. Wang, Z., et al. Spatial association of surface water quality and human cancer in China. *Npj Clean Water* **6**, 53 (2023).
34. Zhang, S. et al. Cancer incidence and mortality in China, 2015. *J. Natl. Cancer Cent.* **1**, 2–11 (2021).
35. Zeng, H. et al. Cancer survival in China, 2003–2005: a population-based study. *Int. J. Cancer* **136**, 1921–1930 (2015).
36. Yang, L. et al. Burden of lung cancer attributable to ambient fine particles and potential benefits from air quality improvements in Beijing, China: A population-based study. *Sci. Total Environ.* **738**, 140313 (2020).
37. Tan, X. et al. A review of current air quality indexes and improvements under the multi-contaminant air pollution exposure. *J. Environ. Manag.* **279**, 111681 (2021).
38. Nong, X., Shao, D., Zhong, H. & Liang, J. Evaluation of water quality in the South-to-North Water Diversion Project of China using the water quality index (WQI) method. *Water Res* **178**, 115781 (2020).
39. Davalos, A. D., Luben, T. J., Herring, A. H. & Sacks, J. D. Current approaches used in epidemiologic studies to examine short-term multipollutant air pollution exposures. *Ann. Epidemiol.* **27**, 145–153 e141 (2017).
40. Anselin, L. J. G. A. Local Indicators of Spatial Association—LISA. *Geogr. Anal.* **27**, 93–115 (1995).
41. Environmental Planning Institute, Ministry of Ecology and Environment. Report on Air Pollution Control Strategy of the 14th Five-Year Plan in Beijing-Tianjin-Hebei and Surrounding Areas (2021).
42. Yang G., Zhuang D. *Atlas of the Huai River Basin Water Environment: Digestive Cancer Mortality*. 1 edn, Springer Dordrecht (2014).
43. Ministry of Ecology and Environment of China. Report on the State of the Ecology and Environment in China. (2021).
44. Paget, S. The distribution of secondary growths in cancer of the breast. *Lancet* **133**, 571–573 (1889).
45. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4765–4774 (2017).
46. Shi, J. F. et al. Is it possible to halve the incidence of liver cancer in China by 2050? *Int J. Cancer* **148**, 1051–1065 (2021).
47. Datzmann, T. et al. Outdoor air pollution, green space, and cancer incidence in Saxony: a semi-individual cohort study. *BMC Public Health* **18**, 715 (2018).
48. Raaschou-Nielsen, O., Ketzel, M., Harbo Poulsen, A. & Sørensen, M. Traffic-related air pollution and risk for leukaemia of an adult population. *Int J. Cancer* **138**, 1111–1117 (2016).

49. Ministry of Ecology and Environment of China. Water quality-Determination of permanganate index, GB11892-89 (1989).

50. Picetti, R. et al. Nitrate and nitrite contamination in drinking water and cancer risk: A systematic review with meta-analysis. *Environ. Res.* **210**, 112988 (2022).

51. Dixon, K. & Kopras, E. Genetic alterations and DNA repair in human carcinogenesis. *Semin. Cancer Biol.* **14**, 441–448 (2004).

52. Hendryx, M., Conley, J., Fedorko, E., Luo, J. & Armistead, M. Permitted water pollution discharges and population cancer and non-cancer mortality: toxicity weights and upstream discharge effects in US rural-urban areas. *Int. J. Health Geogr.* **11**, 9 (2012).

53. Ministry of Ecology and Environment of the People's Republic of China. Technical regulation for selection of ambient air quality monitoring stations. (2013).

54. Ministry of Ecology and Environment of the People's Republic of China. Technical Specifications for Surface Water Environmental Monitoring. (2022).

55. Ministry of Ecology and Environment of the People's Republic of China. Technical Specification for Surface Water Quality Assessment. (2022).

56. National Bureau of Statistics of China. China Statistical Yearbook (2020).

57. Xu X. Multi-annual district and county administrative boundary data of China. Resource and environmental science data registration and publication system. (2023).

58. Hilbe, J. M *Negative Binomial Regression.* (Cambridge University Press, 2011).

59. Mansournia, M. A. & Altman, D. G. Population attributable fraction. *Br. Med. J.* **360**, k757 (2018).

60. Ezzati, M. et al. Estimates of global and regional potential health gains from reducing multiple major risk factors. *Lancet* **362**, 271–280 (2003).

Acknowledgements

Author contributions

J.J. and C.J. contributed to the study's conception and design. L.Z., W.G., C.Y. and Y.S. contributed to data analysis. J.J., L.Z., Z.W., W.G., C.Y. and Y.S. contributed to data interpretation. J.Z., W.H., Y.H. and W.C. contributed to literature research. F.X., X.G. and H.L. contributed to data collection. P.W., Y.C., Y.Z. and J.D. contributed to figure design. J.J., L.Z., Z.W., W.G., C.Y. and Y.S. wrote the manuscript. All authors had access to the raw data, gave critical revisions for important intellectual content and gave final approval of the version to be published.

Competing interests

The authors declare no competing interests.

Additional information

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