Supplementary Information for

Health benefits of on-road transportation pollution control programs in China

Haikun Wang[a,b,1,*], Xiaojing He[b,1], Xinyu Liang[c,1], Ernani F. Choma[d,e], Yifan Liu[b], Li Shan[b], Haotian Zheng[c,f], Shaojun Zhang[c,*], Chris P. Nielsen[f], Shuxiao Wang[c], Ye Wu[c,*], John S. Evans[d]

[a.] Joint International Research Laboratory of Atmospheric and Earth System Sciences, School of Atmospheric Sciences, Nanjing University, Nanjing 210023, P.R. China

[b.] State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing 210023, P.R. China

[c.] School of Environment and State Key Joint Laboratory of Environment Simulation and Pollution Control, Tsinghua University, Beijing 100084, P.R. China

[d.] Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA 02115, USA

[e.] Population Health Sciences, Harvard University, Boston, MA 02115, USA

[f.] Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

[1] These authors contributed equally to this work.

*Corresponding authors:

Haikun Wang, Email: wanghk@nju.edu.cn; Ye Wu, Email: ywu@mail.tsinghua.edu.cn; Shaojun Zhang, Email: zhsjun@tsinghua.edu.cn

www.pnas.org/cgi/doi/10.1073/pnas.1921271117

Fig. S1 Framework of the integrated assessment in this study.

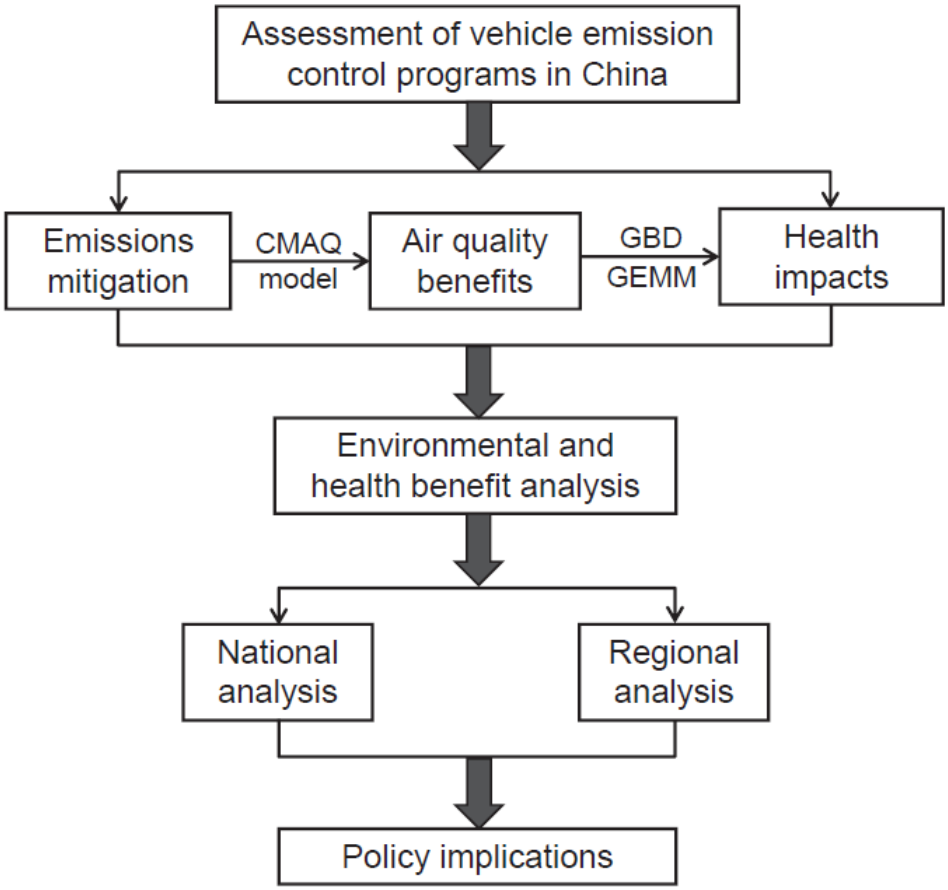


Fig. S2 China's annual actual emissions (*i.e.*, with controls) from on-road vehicles (blue) and the impacts of vehicle control programs on emissions (orange) from 1998 to 2015. a) NMVOCs, b) CO, c) NO_X, and d) $PM_{2.5}$.

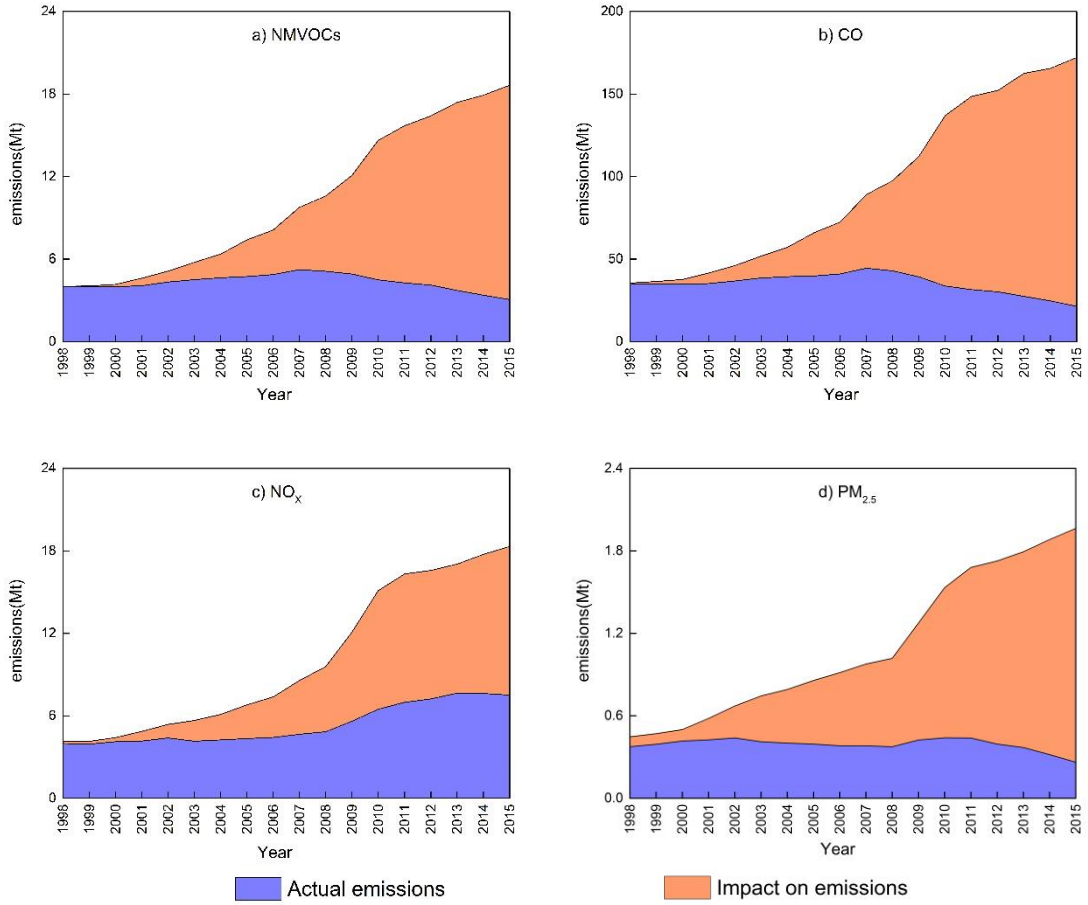


Fig. S3 Decomposition of the changes in vehicular emissions of CO, NMVOCs and $PM_{2.5}$ during the periods of 2005-2010 and 2010-2015, respectively. a) CO, b) NMVOCs, and c) $PM_{2.5}$. The gray bars represent the annual actual vehicle emissions (*i.e.*, with controls). The dark colored bars on the left of dashed line represent the impacts of vehicle kilometers traveled (VKT) by various vehicle fleets on driving the emissions growth during each period. And the light colored bars on the right of dashed line represent the impacts of vehicle control programs on driving the emissions decline. Vehicle fleets include heavy-duty trucks (HDTs), light-duty truck (LDT), medium-duty passenger vehicle (MDPV), heavy-duty passenger vehicle (HDPV), light-duty passenger vehicle (LDPV), motorcycle, taxi and bus.

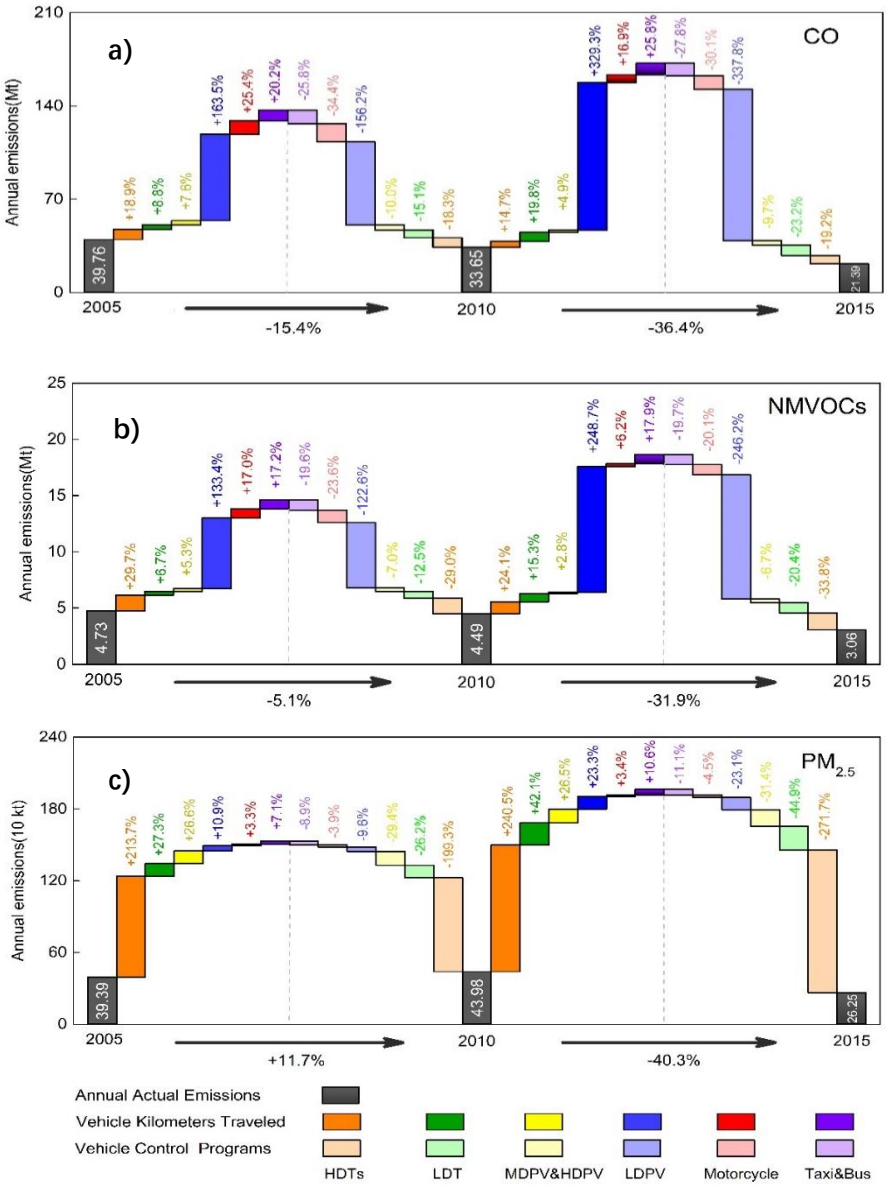


Fig. S4 The proportions of vehicular emissions of NMVOCs, CO, $PM_{2.5}$ and NO_X lower than they would have been without controls in 2005, 2010 and 2015.

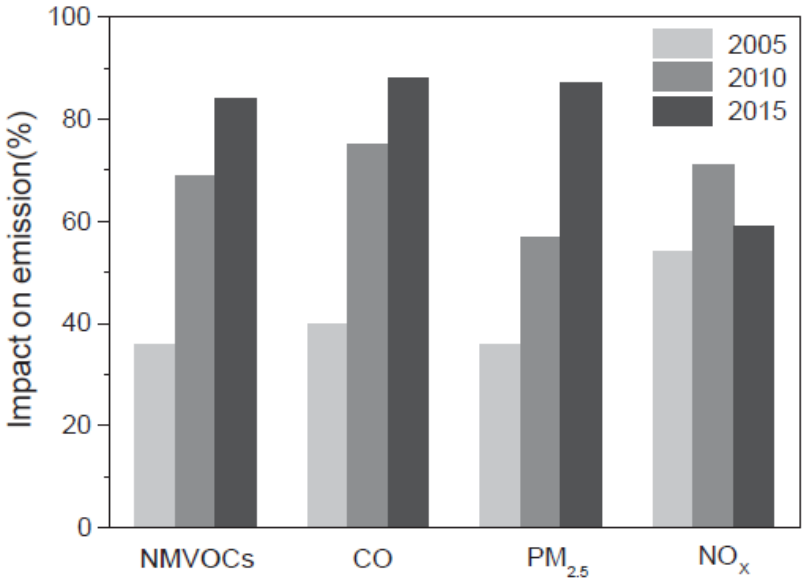


Fig. S5 Impacts of vehicle control programs since 1998 on ambient a) $PM_{2.5}$ and b) O_3 concentrations across China in 2005, 2010 and 2015. Black and red columns represent the population weighted and spatial average of annual concentrations, respectively. The data of impacts represent the projected difference between with and without control scenarios. Positive values imply the actual (*i.e.*, with controls) annual average $PM_{2.5}$ (or O_3) concentrations in that year are lower than it would have been without controls, and vice versa.

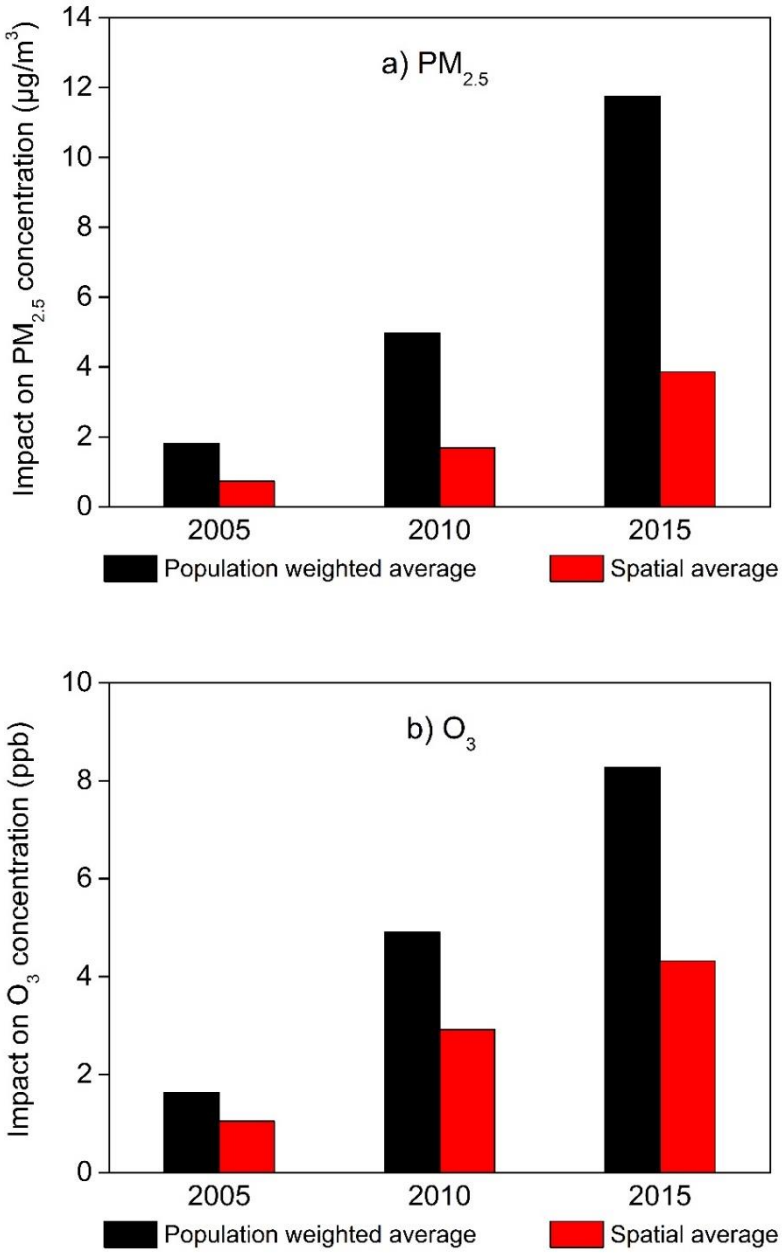


Fig. S6 Geographic distribution of the impacts of vehicle control programs since 1998 on annual average $PM_{2.5}$ concentrations (left column) and related deaths (right column) in 2005, 2010 and 2015. The data of impacts represent the projected difference between with and without control scenarios. Positive values imply the actual (*i.e.*, with controls) annual average $PM_{2.5}$ concentrations and deaths in that year are lower than they would have been without controls, and vice versa.

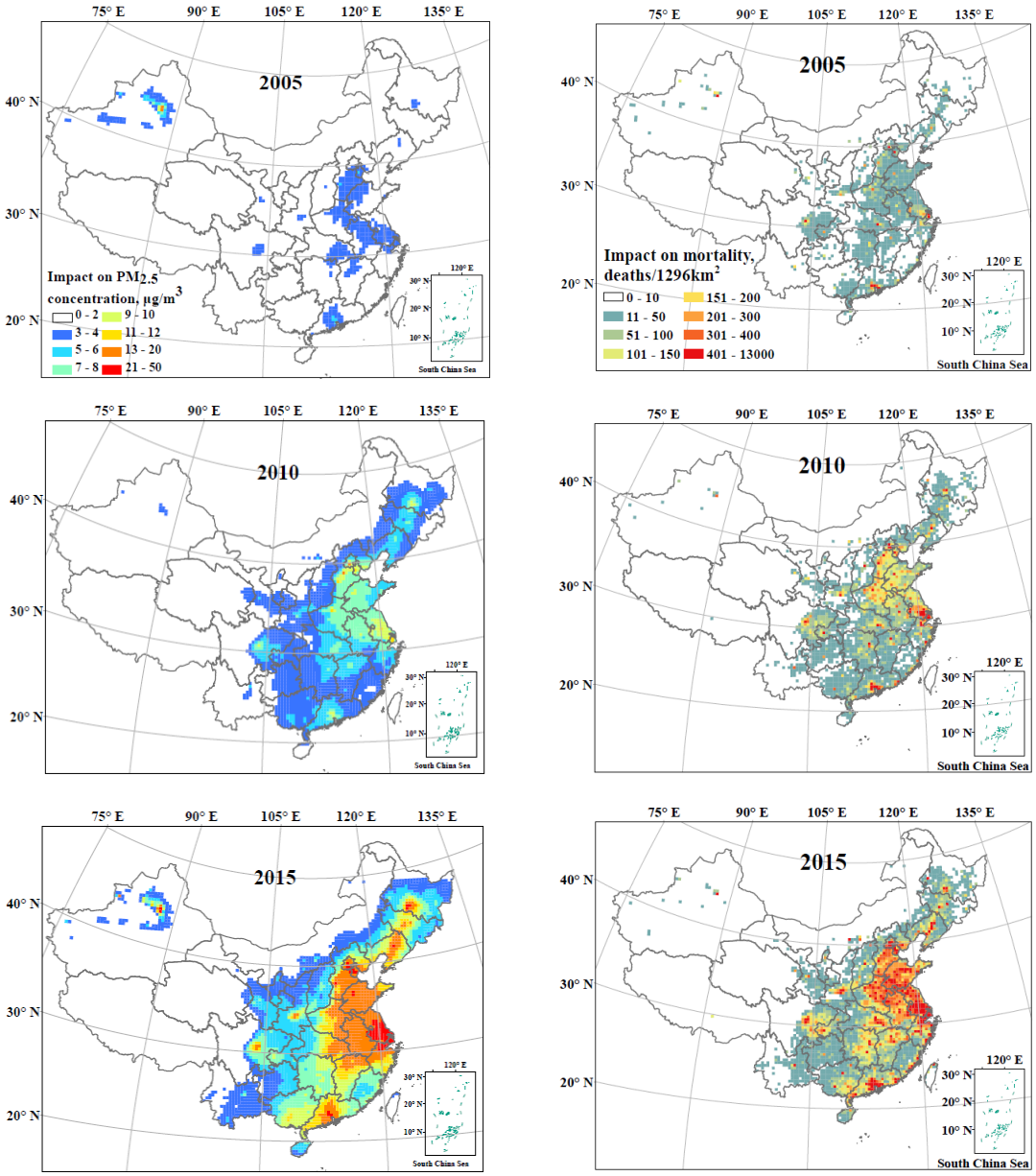


Fig. S7 **The actual (*i.e.*, with controls) vehicular emissions (orange columns), the projected impacts of vehicle control programs on mitigating emissions (*i.e.*, difference in emissions between with and without controls) of NMVOCs, $PM_{2.5}$, NO_X and CO (blue columns), and the corresponding reduction proportion (points) for various provinces of China in 2015.**

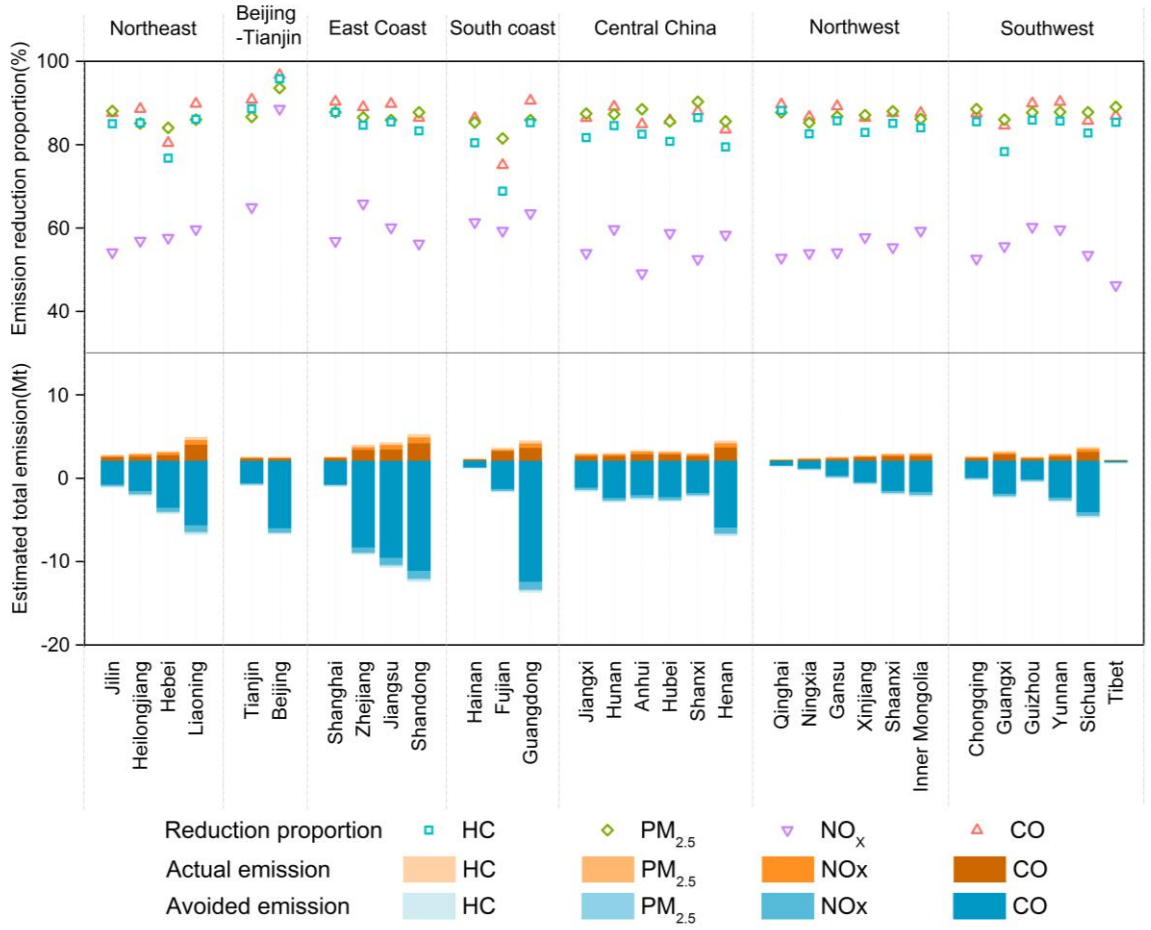


Fig. S8 Geographic distribution of the impacts of vehicle control programs since 1998 on seasonal (April-September) average 8h maximum O_3 concentrations (left) and related deaths (right) in 2005, 2010 and 2015. The data of impacts represent the projected difference between with and without control scenarios. Positive values imply the actual (*i.e.*, with controls) O_3 concentrations and deaths in that year are lower than they would have been without controls, and vice versa.

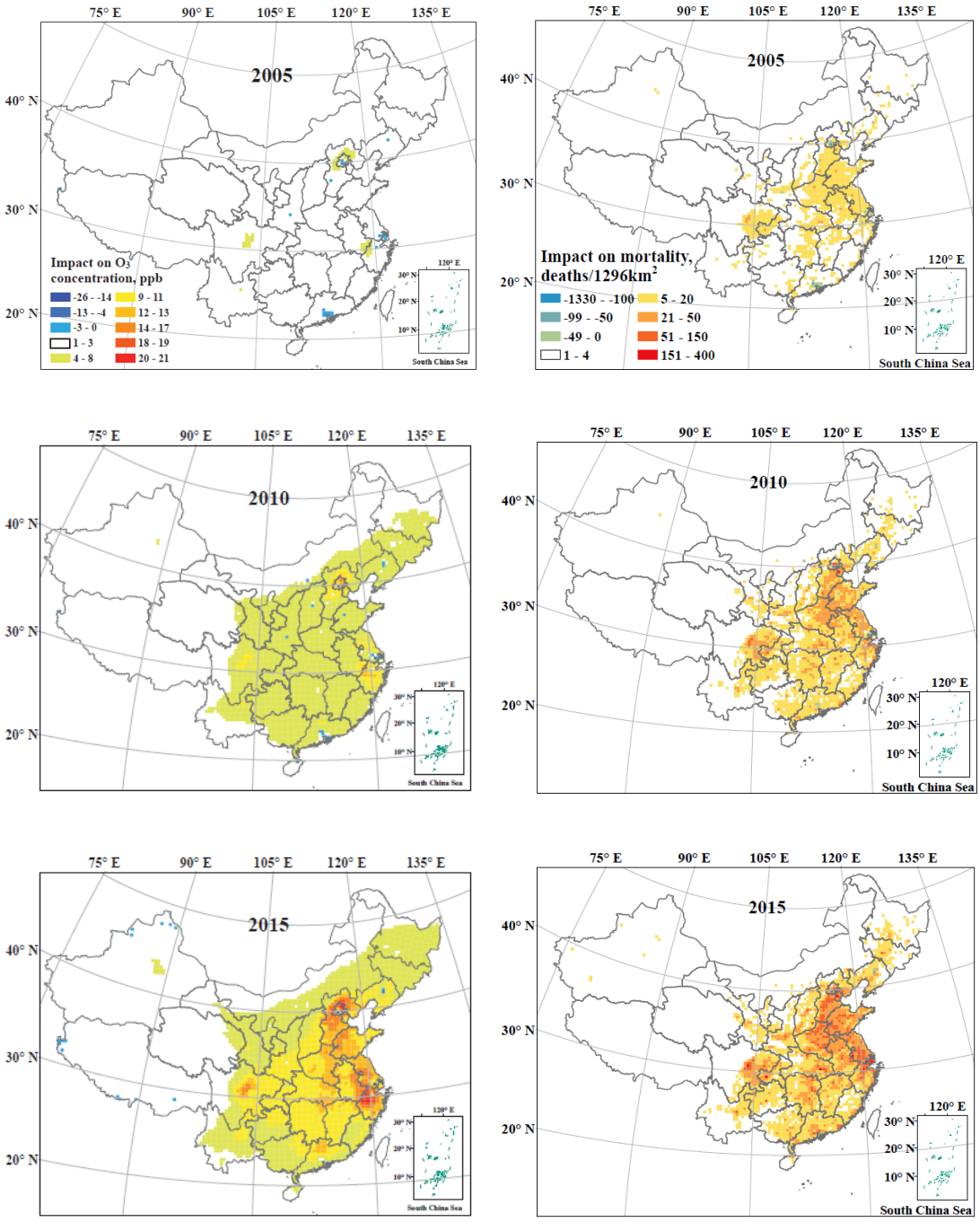


Table S1 **Impacts of vehicle control programs since 1998 on annual average $PM_{2.5}$ concentrations of various provinces in 2005, 2010 and 2015.**

Province	Spatial average ($\mu g/m^3$)			Population weighted average ($\mu g/m^3$)		
	2005	2010	2015	2005	2010	2015
Jilin	0.98	2.99	6.82	1.27	3.96	8.81
Heilongjiang	0.54	1.73	3.53	1.21	3.79	8.05
Hebei	1.57	5.12	10.22	2.17	7.27	12.52
Liaoning	1.28	3.85	10.44	1.63	4.53	12.45
Tianjin	2.72	8.05	18.31	3.51	9.33	23.26
Beijing	1.73	6.06	14.19	3.17	11.13	33.44
Shanghai	3.00	8.57	27.14	3.64	10.15	28.17
Zhejiang	1.42	4.25	14.51	1.73	5.40	17.36
Jiangsu	2.16	7.52	22.05	2.33	8.12	23.88
Shandong	1.93	6.29	15.22	1.97	6.59	15.80
Hainan	0.53	0.85	2.83	1.11	1.57	5.09
Fujian	0.92	2.44	7.14	1.03	2.81	7.78
Guangdong	1.72	3.78	10.24	2.58	4.98	14.11
Jiangxi	1.65	3.99	11.04	1.88	4.69	11.52
Hunan	1.59	3.70	8.12	1.73	4.09	7.93
Anhui	1.99	5.85	17.44	2.12	6.31	17.17
Hubei	1.51	3.95	9.36	2.03	4.89	11.40
Shanxi	1.07	2.65	5.13	1.36	3.23	5.26
Henan	1.77	5.60	11.22	2.03	6.54	11.84
Qinghai	0.18	0.37	0.43	0.93	2.36	2.18
Ningxia	0.81	1.80	3.08	0.84	2.09	3.65
Gansu	0.57	1.14	1.89	1.11	2.20	3.51
Xinjiang	0.72	0.37	0.64	4.12	1.16	1.78
Shaanxi	0.94	2.29	4.94	1.46	3.47	7.46
Inner Mongolia	0.27	0.76	1.46	0.62	1.72	3.33
Chongqing	1.08	3.09	6.79	1.27	3.51	7.21
Guangxi	1.28	3.11	7.36	1.44	3.56	8.47
Guizhou	1.00	2.28	4.64	0.95	2.19	4.01
Yunnan	0.44	0.96	1.74	0.61	1.31	2.46
Sichuan	0.57	1.56	2.65	1.54	4.19	6.21
Tibet	0.02	-0.05	0.00	0.03	0.02	0.30
Total	0.73	1.69	3.86	1.82	4.97	11.75

Table S2 Impacts of vehicle control programs since 1998 on mortality attributable to ambient $PM_{2.5}$ of various provinces in 2005, 2010 and 2015.

Province	2005	2010	2015
Jilin	921	2,978	6,867
Heilongjiang	1,257	4,077	8,843
Hebei	3,548	12,471	23,952
Liaoning	1,851	5,477	15,365
Tianjin	833	2,694	8,189
Beijing	1,178	5,197	17,129
Shanghai	1,678	5,930	18,423
Zhejiang	2,305	8,433	31,466
Jiangsu	4,218	15,588	47,858
Shandong	4,342	14,869	40,673
Hainan	259	435	1,608
Fujian	1,119	3,367	10,843
Guangdong	6,264	14,542	51,479
Jiangxi	2,097	5,706	17,287
Hunan	2,764	7,152	16,627
Anhui	3,148	9,322	27,183
Hubei	2,798	7,059	18,871
Shanxi	1,312	3,512	7,077
Henan	4,599	15,224	31,337
Qinghai	152	413	505
Ningxia	139	383	793
Gansu	780	1,669	3,053
Xinjiang	2,878	979	1,878
Shaanxi	1,420	3,633	8,399
Inner Mongolia	523	1,423	3,206
Chongqing	955	2,800	6,602
Guangxi	1,755	4,529	12,695
Guizhou	933	2,144	4,353
Yunnan	817	1,915	4,378
Sichuan	3,187	8,987	15,604
Tibet	5	24	136
Total	60,035	172,932	462,679

Table S3 Impacts of vehicle control programs since 1998 on seasonal (April–September) 8h maximum O_3 concentrations (spatial average) and related mortality of various provinces in 2005, 2010 and 2015.

Province	Impact on O_3 concentration (ppb)			Impact on mortality (COPD)		
	2005	2010	2015	2005	2010	2015
Jilin	1.48	3.87	6.11	211	438	743
Heilongjiang	0.75	1.99	3.02	201	414	645
Hebei	2.30	7.00	11.57	762	2,123	3,741
Liaoning	1.70	4.66	7.56	296	686	955
Tianjin	2.13	8.38	12.30	90	449	594
Beijing	2.58	10.07	14.15	-270	-101	-955
Shanghai	0.88	3.64	10.15	-55	-11	853
Zhejiang	1.83	7.04	11.45	402	1,624	3,009
Jiangsu	1.97	6.12	11.57	621	1,611	3,454
Shandong	2.10	6.17	11.13	973	2,499	4,697
Hainan	0.34	0.55	1.11	29	-2	90
Fujian	1.59	5.11	7.05	215	596	850
Guangdong	1.49	4.74	7.01	225	1,218	2,171
Jiangxi	2.18	6.64	9.75	458	1,287	1,922
Hunan	2.01	6.23	9.70	631	1,747	2,738
Anhui	2.41	7.24	12.03	696	1,800	2,918
Hubei	2.16	6.52	10.21	602	1,591	2,412
Shanxi	1.97	5.37	8.95	313	753	1,397
Henan	2.20	6.39	10.99	997	2,567	4,395
Qinghai	0.50	1.21	1.36	32	71	107
Ningxia	1.44	4.06	5.86	39	92	141
Gansu	1.22	3.23	4.48	215	530	725
Xinjiang	0.56	0.94	1.19	81	102	160
Shaanxi	1.86	5.47	8.03	305	780	1,239
Inner	0.76	2.05	3.06	133	286	476
Chongqing	2.21	6.52	9.23	322	806	1,102
Guangxi	1.46	4.30	6.48	346	896	1,306
Guizhou	1.72	5.29	7.64	297	789	1,091
Yunnan	1.12	3.01	4.05	316	746	1,038
Sichuan	1.65	4.76	6.36	972	2,348	3,290
Tibet	0.56	0.94	1.19	6	12	8
Total	1.05	2.92	4.32	10,461	28,747	47,312

Table S4 Impacts of vehicle control programs since 1998 on mortality attributable to ambient $PM_{2.5}$ and O_3 of various provinces in 2005, 2010 and 2015.

Province	2005	2010	2015
Jilin	1132	3416	7610
Heilongjiang	1458	4491	9488
Hebei	4310	14594	27693
Liaoning	2147	6163	16320
Tianjin	923	3143	8783
Beijing	908	5096	16174
Shanghai	1623	5919	19276
Zhejiang	2707	10057	34475
Jiangsu	4839	17199	51312
Shandong	5315	17368	45370
Hainan	288	433	1698
Fujian	1334	3963	11693
Guangdong	6489	15760	53650
Jiangxi	2555	6993	19209
Hunan	3395	8899	19365
Anhui	3844	11122	30101
Hubei	3400	8650	21283
Shanxi	1625	4265	8474
Henan	5596	17791	35732
Qinghai	184	484	612
Ningxia	178	475	934
Gansu	995	2199	3778
Xinjiang	2959	1081	2038
Shaanxi	1725	4413	9638
Inner Mongolia	656	1709	3682
Chongqing	1277	3606	7704
Guangxi	2101	5425	14001
Guizhou	1230	2933	5444
Yunnan	1133	2661	5416
Sichuan	4159	11335	18894
Tibet	11	36	144
Total	70496	201679	509991

Table S5 Parameters used to describe the overall shape of the concentration-response relationship in the GEMM LRI+NCD (Burnett *et al.*, 2018).

Age Group (years)	θ_j	Standard Error θ_j	α_j	μ_j	ν_j
Adults (>25)	0.143	0.01807	1.6	15.5	36.8
25-29	0.1585	0.01477	1.6	15.5	36.8
30-34	0.1577	0.01470	1.6	15.5	36.8
35-39	0.157	0.01463	1.6	15.5	36.8
40-44	0.1558	0.01450	1.6	15.5	36.8
45-49	0.1532	0.01425	1.6	15.5	36.8
50-54	0.1499	0.01394	1.6	15.5	36.8
55-59	0.1462	0.01361	1.6	15.5	36.8
60-64	0.1421	0.01325	1.6	15.5	36.8
65-69	0.1374	0.01284	1.6	15.5	36.8
70-74	0.1319	0.01234	1.6	15.5	36.8
75-79	0.1253	0.01174	1.6	15.5	36.8
80+	0.1141	0.01071	1.6	15.5	36.8

Table S6 The cessation lag of impacts of vehicle control programs since 1998 on mortality attributable to 2005, 2010 and 2015 ambient $PM_{2.5}$ pollution, respectively. We applied the lag structure recommended by the US EPA Science Advisory Board and used in its periodic analysis of Section 812 of the Clean Air Act (Walton, 2010).

Lag/Year	The first Year		
	2005	2010	2015
1	18,011	51,879	138,803
2	7,505	21,616	57,835
3	7,505	21,616	57,835
4	7,505	21,616	57,835
5	7,505	21,616	57,835
6	800	2,306	6,169
7	800	2,306	6,169
8	800	2,306	6,169
9	800	2,306	6,169
10	800	2,306	6,169
11	800	2,306	6,169
12	800	2,306	6,169
13	800	2,306	6,169
14	800	2,306	6,169
15	800	2,306	6,169
16	800	2,306	6,169
17	800	2,306	6,169
18	800	2,306	6,169
19	800	2,306	6,169
20	800	2,306	6,169
Total	60,035	172,932	462,679

Note S1. Scenario vehicular emissions

We developed detailed emission factors based on vehicle type, emission standards, vehicle age, fuel type, and calendar year, to analyze on-road vehicle emissions in China (Wu et al., 2016). We developed survival rate functions based on millions of vehicle registration and scrappage records from both Tier-1 (*e.g.*, Beijing, Shanghai) and internal cities in China, which could facilitate a continuous estimation of vehicle age distribution.

For light-duty gasoline vehicles (LDPVs; *i.e.*, passenger cars), the emission factors were measured based on dynamometer tests of more than one thousand vehicles (Zhang et al., 2014), and on a few recent updates based on the measurements of China 5 emission standard vehicles. A series of corrections of real-world fuel use conditions, such as driving speed, sulfur content of gasoline fuels, travel distance that affect evaporative and cold start emissions, have been included. For heavy-duty diesel vehicles, the basic emission factors and major correction factors (*e.g.*, vehicle weight, driving speed, and auxiliary load) were developed based on portable emission measurement systems (Zhang et al., 2014; Wu et al., 2012). As for vehicle mileage, we consulted with the environmental authority and obtained the fleet-average annual vehicle kilometers traveled (VKT) data from official inspection records. We observed a decline in the fleet-average annual VKT values of LDPVs, since a large fraction of LDPVs were used for commercial purposes around 2000 but since 2010 private use has been predominant. For heavy-duty diesel trucks (HDDTs), the fleet-average annual VKT values increased in this period due to the enhanced regional and cross-regional freight connectivity that required more long-distance travels. See details in Dataset S1A and B for the key parameters used to estimate vehicular emissions in this study.

As shown in Table 1 of the main-text, we designed a “without (w/o) control” scenario to represent the emission trends without vehicular emission control programs in China, which was developed in our previous study (Wu et al., 2016). Of note, we assumed that no vehicle emission controls would be implemented under the w/o scenario except for the natural fleet turnover. For example, under the w/o control scenario, in 2015 all new vehicles comply with pre-China 1 standards and the fuel quality is set according to the

actual level in 1998. Furthermore, subsidized scrappage, promotion of alternative fuel vehicles, and enhanced inspection/maintenance (I/M) program, would all be absent under the w/o control scenario. We estimate the impacts of vehicle control programs on annual emissions as the gap between estimates under with and w/o control scenarios from 1998 to 2015.

The vehicle emissions model is a statistical model based on measurement data. Therefore, the uncertainty in emission factors and total emissions could be estimated based on the probability distribution functions of all key sensitive parameters. In Zhang et al. (2014) and Wu et al. (2016), we have built probability-based distribution functions to address the uncertainties of key parameters of the emissions model based on detailed experimental data and investigation results. Uncertainties in vehicle emission factors and in the total emissions were analyzed with the Monte Carlo simulations by taking account of the probability distributions of these key model parameters or input variables. Uncertainty ranges of vehicular emissions at a 95% confidence level (CI 95) were estimated for each year by running 100 thousand trials of Monte Carlo simulations. The uncertainty information of China's vehicular emissions under the with-control scenarios has been updated to 2015 in this study and is shown in the Dataset S1C. For the emissions under the w/o control scenario, we referred to the uncertainty ranges (Cl 95) estimate for the year of 1998 (*i.e.*, before the implementation of vehicle emission controls) under the with control scenario.

Note S2. Air pollution exposures modelling

CMAQ simulating. A one-way, triple-nesting method in the Weather Research and Forecasting (WRF) model and Community Multiscale Air Quality (CMAQ) v 5.0.1 model was used to simulate the air pollutant concentrations for scenarios with and w/o policy intervention. Domain 1 at a grid resolution of 36 km × 36 km covers the Greater China region, including Mainland China and part of East Asia and Southeast Asia regions; Domain 2 covers the Beijing-Tianjin-Hebei (BTH) and Yangtze River Delta (YRD) regions as a finer-scale case study at a resolution of 4 km × 4 km.

Consistent with our previous studies (Wang et al., 2011; Zhao et al., 2016; Liang et al., 2019), we selected WRF simulations for four representative months (January, May, August and November) as the meteorological input for the CMAQ simulations. We incorporated a two-dimensional volatility basis set (2D-VBS) technique in the CMAQ model to improve the secondary organic aerosol (SOA) simulation, because the default SOA chemistry used in the CMAQ v 5.0.1 significantly underestimated SOA concentrations in China (Zhao et al., 2016).

Different from the vehicle emissions model, the CMAQ model is an atmospheric physics and chemistry-based tool (not a statistical model), and the simulation performance is influenced by a series of factors, primarily including the physics and chemistry mechanisms applied in the CMAQ model, the accuracy of emission inventory data, and the representativeness of meteorological simulations. Previous papers have illustrated the detailed model setup (Liang et al., 2019) and validation against field observation results (Zhao et al., 2013; 2016). For example, Zhao et al. (2013) verified the good agreements of simulated hourly $PM_{2.5}$ concentrations in Beijing and Shanghai against the concurrent observations with the normalized mean bias (NMBs) of -8.5% and -13%, respectively, based on cross-season comparisons in May, August and November. We recently evaluated the national-scale simulation performance of $PM_{2.5}$ in 2015 by comparing against the ground-level observation data. The monthly NMBs are estimated to range from -2% to -34%, with an annual NMB of approximately -10%. Slight underpredictions tend to be found in wintertime pollution episodes with rapid growth

of $PM_{2.5}$ mass concentrations, and the photochemistry in the wintertime production of secondary aerosol is considered to be one important cause of the underrated $PM_{2.5}$ concentrations (Fu et al., 2020).

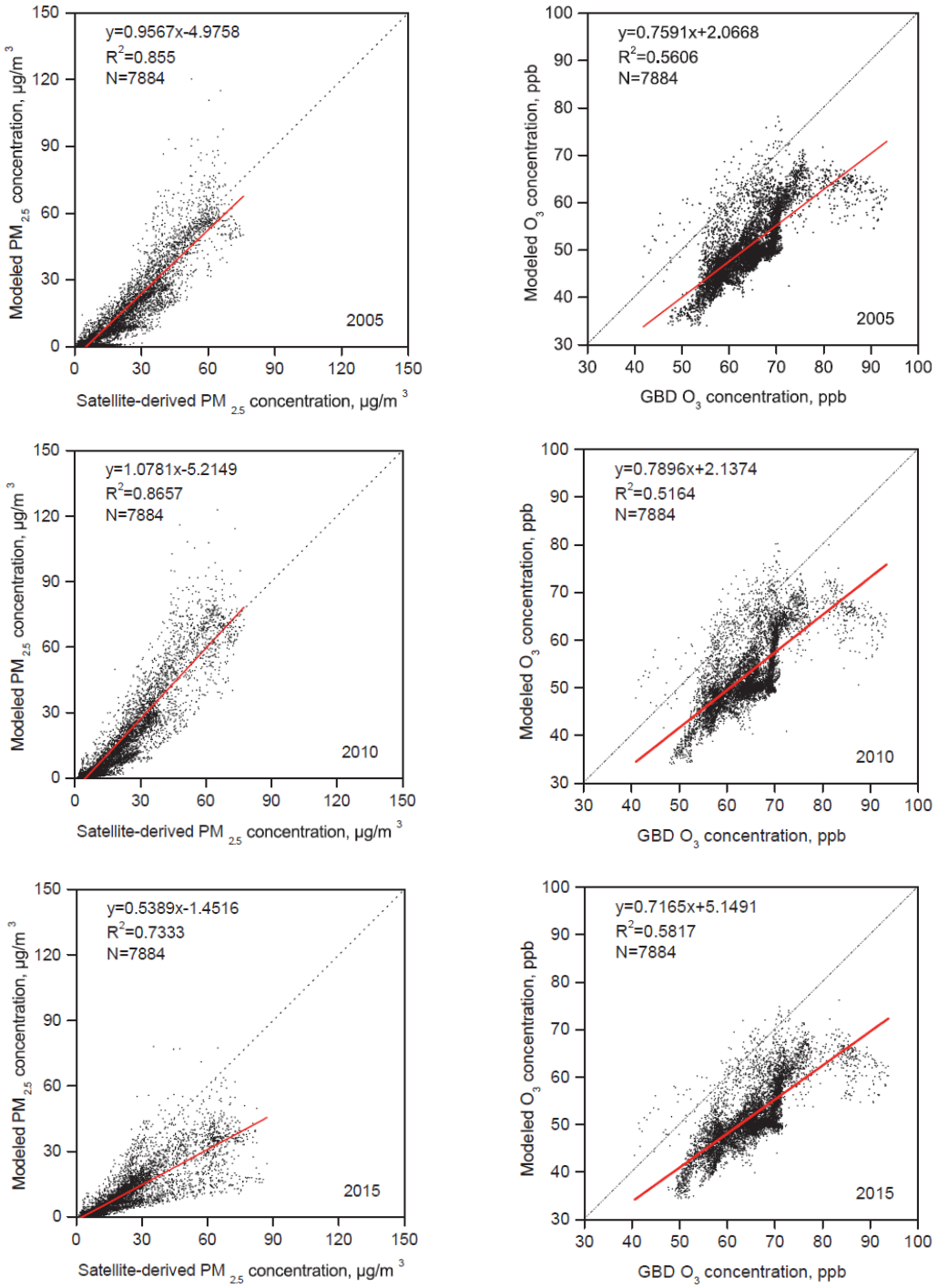
For O_3, both the air quality simulation and the health impact assessment focus on the high-concentration period (e.g., warm season). Ding et al. (2019) evaluated the CMAQ simulation results of warm-season O_3 concentrations (from April to September) against the observation results released by the Ministry of Ecological Environment. The comparison metric was daily maximum 8-h averages (MDA) across China (grid resolution of 27 km × 36 km). In general, the daily-specific validations suggest most of the comparison dates were within the range of 1:2 line to 2:1 line. The validation results for 2015 following the same method show that the monthly NMBs range from +2% to +19% with an average of +9.8%.

Calibrating of simulated concentrations. Different from many other studies simply relying on air quality simulation results (e.g., Liang et al., 2019), we further utilized satellite retrieved and other observation data to correct the potential simulation bias (see SI Appendix, Fig. S9) caused by the imperfect meteorological data and/or incomplete atmospheric chemical reaction schemes (US EPA, 2007; Wang et al., 2017; Zhang et al., 2017). This approach can well address the spatial heterogeneity of the bias between air quality simulations and observed results across a vast country.

We used the modeled relative changes in $PM_{2.5}$ concentrations between with and without control scenarios (as defined in Table 1 in the main-text), and multiplied them to surface concentrations derived from satellite observations (van Donkelaar et al., 2015; 2016) (see Equation 2 and 3 in the main-text) to estimate the impact of vehicle control programs since 1998 on ambient $PM_{2.5}$ concentrations. According to Boys et al. (2014) and van Donkelaar et al. (2015), the uncertainty in satellite-derived $PM_{2.5}$ decreases with an increase in sampling days, and annual mean $PM_{2.5}$ satellite retrievals are estimated using a 3-year moving average. This satellite-derived $PM_{2.5}$ concentrations dataset has already been calibrated by surface observations, with a small uncertainty of approximately ±5% on average (Brauer et al., 2016). It has been applied broadly in the literature, including notably in the Global Burden of Disease (GBD) assessments, to

represent the spatiotemporal distribution of $PM_{2.5}$ exposures. Following a similar procedure to $PM_{2.5}$, we applied GBD's O_3 concentrations data (Stanaway et al., 2018) to calibrate the CMAQ simulated impacts on ambient O_3 pollution.

Fig. S9 Comparisons of CMAQ simulated results with satellite-derived $PM_{2.5}$ concentrations (left) and O_3 concentrations used in GBD 2017 (right) for 2005, 2010 and 2015.



Note S3. Heath effects attributed to air pollution

S3.1. Chronic and acute health effect

Particulate matter can cause, in addition to mortality, exacerbation of cardiovascular and respiratory disease, such as asthma, and decrease in lung function (Bell and Samet, 2010). Previous air pollution research has shown that the absolute frequency of non-fatal outcomes is much larger than that of fatal outcomes, but the latter represents the vast majority of monetized impacts. A study of the 2013 severe haze event in the Beijing area estimated that 74% of the monetized impacts of $PM_{2.5}$ were due to 690 attributable deaths, despite the tens of thousands of asthma and acute bronchitis cases and clinic visits, as well as thousands of hospitalizations (Gao et al., 2015). Similarly, in the United States the EPA estimates that 230,000 annual attributable deaths avoided comprise in excess of 95% of the estimated monetized benefits of the Clean Air Act in the 1990-2020 period, despite the millions of annual cases of asthma and upper and lower respiratory symptoms, among others (EPA, 2011). The EPA study covers both fine particles and ozone.

Previous research has also shown that despite the severe haze episodes in China, the vast majority of deaths attributable to $PM_{2.5}$ are a consequence of chronic exposure (Wang et al., 2020). Wang and colleagues estimate 1.21 million yearly deaths attributable to chronic $PM_{2.5}$ in China in the 2013-2017 period, a number 10 times larger than the estimated 116 thousand attributable to acute $PM_{2.5}$ exposures, showing that chronic exposure is a much larger public health issue in the country. Therefore, as mortality is responsible for the majority of the public health burden and the mortality due to chronic exposure is much higher than that due to acute exposure, we focus our study on mortality attributable to chronic exposure in this study.

S3.2. Comparisons between IER, GEMM and Chinese male cohort

In this study, we applied the E-R function developed by Burnett et al. (2018), known as the Global Exposure Mortality Model (GEMM). GEMM is a recent synthesis of evidence that includes data from 41 of the largest cohorts worldwide; more importantly, it is a collaboration among the investigators responsible for 15 of them, for which they had

access to subject-level data. We use GEMM because it includes recent epidemiological evidence – including especially the only available national level cohort study on outdoor $PM_{2.5}$ pollution in China (*i.e.*, Yin et al., 2017) – that extends the range of $PM_{2.5}$ exposures to concentrations of up to 84 μg m^{-3} in China, crucial to the present study.

Burnett et al. (2018) developed two versions of GEMM: (i) the non-accidental GEMM (GEMM NCD+LRI), which assesses the relationship between $PM_{2.5}$ and non-accidental deaths and which we use; and (ii) the five cause of death GEMM (GEMM 5-COD), which assesses the relationship between $PM_{2.5}$ and five causes of death (ischemic heart disease (IHD), chronic obstructive pulmonary disease (COPD), stroke, acute lower respiratory infection (ALRI) and lung cancer). We primarily use GEMM NCD+LRI because it captures a larger share of deaths attributable to $PM_{2.5}$, when compared to the 5-COD model.

We, however, assess how our results would change if GEMM 5-COD were used instead; as well as if we used another model focusing on the same five causes of deaths: an integrated exposure-response (IER) model (Stanaway et al., 2018) developed by GBD to estimate the health effects attributable to ambient $PM_{2.5}$ exposure. Unlike GEMM, whose epidemiological evidence comes exclusively from studies concerning outdoor $PM_{2.5}$ pollution, the IER incorporated risk information from multiple $PM_{2.5}$ sources (*i.e.*, outdoor air pollution, secondhand smoke, household air pollution from the use of solid fuels, and active smoking). The IER model assumes equivalent exposure and toxicity of $PM_{2.5}$ from multiple sources, and the risk information in the current model comes only from cohort studies conducted in low-polluted Europe and North America. The IER model could be expressed as Eq. (s1):

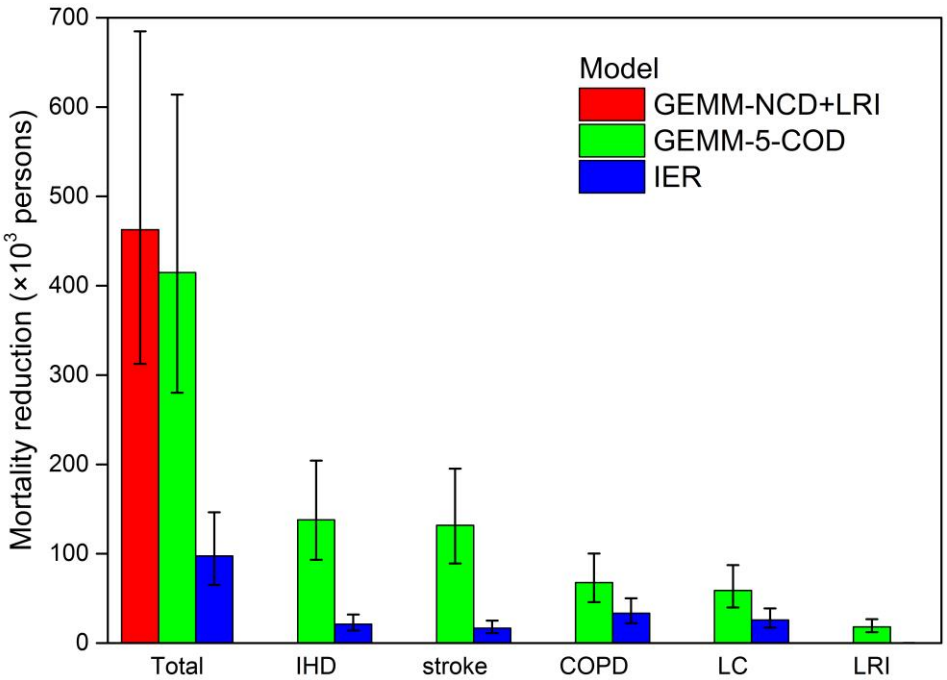
$$RR_j(C_i) = \begin{cases} 1 + \alpha\left[1 - exp(-\gamma(C_i - C_o)^{\delta}\right], & for\ C_i > C_0 \\ 1\ , & for\ C_i \le C_0 \end{cases} \quad \text{(s1)}$$

where $RR_j(C_i)$ is the relative risk of the annual average $PM_{2.5}$ concentration in grid cell *i* for health endpoint *j*; C_o is a theoretical minimum-risk concentration (*i.e.*, 2.4-5.9 μg/m^3) for each health endpoint above which there is evidence indicating health benefits of $PM_{2.5}$ exposure reductions; α, γ and δ are parameters used to describe the overall shape of the concentration-response relationship, which are estimated by GBD

2017 through a stochastic fitting process based on multiple cohort studies worldwide (Stanaway et al., 2018); C_i is the $PM_{2.5}$ concentration in grid cell *i*.

Using GEMM NCD+LRI, GEMM 5-COD, and IER, the impacts of vehicle emissions control programs since 1998 on ambient $PM_{2.5}$ concentrations are estimated to have led to 463 thousand (CI 95: 310 thousand to 680 thousand), 415 thousand (280 thousand to 610 thousand) and 98 thousand (60 thousand to 150 thousand) fewer deaths, respectively, attributable to 2015 ambient $PM_{2.5}$ levels in China (SI Appendix, Fig. S10). Using GEMM NCD+LRI leads to relatively higher estimated health benefits than GEMM 5-COD due to its enhanced statistical power to characterize the shape of the $PM_{2.5}$ mortality associations (Burnett et al., 2018). The difference between the mortality results of GEMM NCD+LRI and IER are much larger, with GEMM NCD+LRI predicting nearly fivefold the health benefits as IER.

Fig. S10 Comparison of the impacts of vehicle control programs on mortality attributable to 2015 ambient $PM_{2.5}$ pollution based on GEMM NCD+LRI, GEMM 5-COD and IER model.



As pointed out by Burnett et al. (2018), GBD might underestimate RR values because the IER used by GBD2017 involved additional sources (not only on outdoor air pollution) of

exposure. GEMM estimates health impacts attributed to outdoor air pollution that are larger than those estimated by the IER model, especially the impacts associated with potential changes in exposures in countries with elevated $PM_{2.5}$ concentrations (*e.g.*, India and China). A recent analysis performed by Xue et al. (2019) also indicates that the census-based results in China were in better consistent with that of GEMM-based than IER-based.

The effect sizes reported by Chinese male cohort study (Yin et al., 2017) are substantially larger than IER's for the range of exposures of interest in our study, and are also larger than GEMM's. For non-accidental deaths, Yin et al. (2017) reports a Hazard Ratio (HR) of 1.09 (1.08-1.09) per 10 $\mu g/m^3$ for their fully adjusted model, that includes both individual- and area-level covariates, and of 1.04 (1.03-1.04) per 10 $\mu g/m^3$ in their model with individual-level covariates only. Applying those to our study[1] and assuming a log-linear form (constant slope), we would have 940,000 (much larger than 510,000 estimated with GEMM) and 410,000 deaths/year in 2015, respectively. GEMM's approach to the results reported by Yin et al. (2017) was to give equal weights to the two models, reporting a HR of 1.064 (1.017-1.115) [Table S1 in Burnett et al. (2018)] for a log-linear form of the E-R function. In our study, that translates to 660,000 (180,000-1,200,000) deaths/year in 2015.

These large mortality estimates, however, might simply reflect the implausibility of a log-linear model for high exposures. As Burnett and Cohen (2020) point out, extending a log-linear form to very high exposures results in biologically implausible estimates of HRs. The $PM_{2.5}$ concentrations in our study exceed the point where the authors indicate log-linear models start to diverge from GEMM (~50 $\mu g/m^3$) as well as the maximum levels in GEMM's epidemiological basis of evidence (84 $\mu g/m^3$). GEMM had access to individual-level data for the Chinese cohort in question (Yin et al., 2017) and could fit their model with a more flexible E-R function, instead of imposing a log-linear form.

[1] Since our counterfactual scenario results in a population-weighted average exposure that is 11.7 $\mu g/m^3$ higher than the actual scenario, this results in [assuming a log-linear form/constant slope] an 11% increase in the 8.8 million NCD+LRI deaths using a HR of 1.09 per 10 $\mu g/m^3$ and a 5% increase with a HR of 1.04 per 10 $\mu g/m^3$.

Finally, the addition of the cohort by Yin et al. (2017) to GEMM's basis of epidemiological evidence has little effect in GEMM's E-R function (Burnett et al., 2018). In view of this and of the lack of knowledge about true heterogeneity, we choose to use a synthesis of evidence which includes that cohort and many others, as opposed to relying on the results of a single study that are subject to considerable uncertainty. The wide 95% CI that ensues from using the combination of models by Yin et al. (2017) (as done by Burnett et al. (2018)) strengthens our argument that the E-R function is the main contributor to the true uncertainty to our results. This is supported by a recent analysis by Burnett and Cohen (2020), who showed the stark differences observed in mortality estimates using IER, GEMM and log-linear models at the high exposures, highlighting the key role of the uncertainty surrounding the E-R function and the critical need for more studies of populations exposed to very high $PM_{2.5}$ concentrations.

References

Bell, M and Samet, J. Chapter 12, Air Pollution, from Frumkin H (editor), Environmental Health: from Global to Local, 2nd Edition, Jossy-Bass Publishers, San Francisco (2010).

Boys, B. L., et al. Fifteen-year global time series of satellite-derived fine particulate matter. Environ. Sci. Technol. 48, 11109−11118 (2014).

Brauer, M., et al. Ambient air pollution exposure estimation for the global burden of disease 2013. Environ. Sci. Technol. 50, 79-88 (2016).

Burnett, R. and Cohen, A. Relative Risk Functions for Estimating Excess Mortality Attributable to Outdoor $PM_{2.5}$ Air Pollution: Evolution and State-of-the-Art. Atmosphere 11, 589 (2020).

Burnett, R., et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proceedings of the National Academy of Sciences of the United States of America 115, 9592-9597 (2018).

Ding, D., et al. Impacts of emissions and meteorological changes on China's ozone pollution in the warm seasons of 2013 and 2017. Front. Environ. Sci. Eng. 13, 76 (2019).

Fu, X., et al. Persistent heavy winter nitrate pollution driven by increased photochemical oxidants in northern China. Environ. Sci. Technol., 54, 3881-3889 (2020).

Gao, M., et al. Health impacts and economic losses assessment of the 2013 severe haze event in Beijing area. Science of the Total Environment 511, 553-561 (2015).

Liang, X., et al. Air quality and health benefits from fleet electrification in China. Nature Sustainability 2, 962-971 (2019).

Stanaway, J. D., et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 392, 1923-1994 (2018).

US EPA (Environmental Protection Agency). The benefits and costs of the Clean Air Act from 1990 to 2020. (2011). Final Report - Rev. A, Available at https://www.epa.gov/sites/production/files/2015-07/documents/fullreport_rev_a.pdf. Accessed March 16, 2020.

US EPA. Letter from Chairs of the Health Effects Sub-Committee and Advisory Council on Clean Air Compliance Analysis to EPA Administrator December 6 2004. (2004). Available at http://yosemite.epa.gov/sab%5CSABPRODUCT.NSF/39F44B098DB49F3C85257170005293E0/$File/council_ltr_05_001.pdf. Accessed March 16, 2020.

US EPA. Guidance on the use of models and other analyses for demonstrating attainment of air quality goals for ozone, $PM_{2.5}$, and regional haze. EPA -454/B-07-002. (2007).

van Donkelaar, A., et al. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. Environ. Sci. Technol. 50, 3762−3772 (2016).

van Donkelaar, A., et al. Use of satellite observations for long-term exposure assessment of

global concentrations of fine particulate matter. Environ. Health Perspect. 123, 135 (2015).

Wang, H., et al. Trade driven relocation of air pollution and health impacts in China. Nature Communications, 8, 738 (2017).

Wang, S., et al. Verification of anthropogenic emissions of China by satellite and ground observations. Atmos. Environ. 45, 6347–6358 (2011).

Walton HA. Development of proposals for cessation lag(s) for use in total impact calculations. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/304655/COMEAP_development_of_proposals_for_cessation_lags.pdf. (2010). Accessed July 3, 2020.

Wu, X., et al. Assessment of vehicle emission programs in China during 1998–2013: Achievement, challenges and implications. Environmental Pollution 214, 556-567 (2016).

Wu, Y., et al. The challenge to NOx emission control for heavy-duty diesel vehicles in China. Atmospheric Chemistry and Physics 12, 9365-9365 (2012).

Xue, T., et al. Change in the number of $PM_{2.5}$-attributed deaths in China from 2000 to 2010: Comparison between estimations from census-based epidemiology and pre-established exposure-response functions. Environment int. 129, 430-437 (2019).

Yin, P., et al. Long-term fine particulate matter exposure and nonaccidental and cause-specific mortality in a large national cohort of Chinese men. Environ. Health Persp. 125, 117002 (2017).

Zhang, Q., et al. Transboundary health impacts of transported global air pollution and international trade. Nature 543, 705–709 (2017).

Zhang, S., et al. Historic and future trends of vehicle emissions in Beijing, 1998–2020: A policy assessment for the most stringent vehicle emission control program in China. Atmospheric Environment 89, 216-229 (2014).

Zhao, B., et al. Impact of national NOx and SO_2 control policies on particulate matter pollution in China. Atmos. Environ. 77, 453–463 (2013).

Zhao, B., et al. Quantifying the effect of organic aerosol aging and intermediate-volatility emissions on regional-scale aerosol pollution in China. Sci. Rep. 6, 28815 (2016).