Long-term air pollution exposure and incident dementia in American elderly population: a national cohort study (2000-2018)

Liuhua Shi (liuhua.shi@emory.edu)
Emory University https://orcid.org/0000-0001-8165-4644
Kyle Steenland
Emory University
Haomin Li
Emory University
Pengfei Liu
Georgia Institute of Technology
Yuhan Zhang
Emory University
Robert Lyles
Emory University
Weeberb Requia
Fundação Getúlio Vargas https://orcid.org/0000-0002-7564-3364
Sindana Ilango
University of Washington
Howard Chang
Emory University
Thomas Wingo
Emory University https://orcid.org/0000-0002-7679-6282
Rodney Weber
Georgia Institute of Technology
Joel Schwartz
Harvard University https://orcid.org/0000-0001-6168-378X

Article

Keywords:

Posted Date: May 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-514522/v1

Version of Record: A version of this preprint was published at Nature Communications on November 19th, 2021. See the published version at https://doi.org/10.1038/s41467-021-27049-2.

Long-term air pollution exposure and incident dementia in an American elderly population: a national cohort study (2000-2018)

Liuhua Shi*[#1], Kyle Steenland[#1], Haomin Li[2], Pengfei Liu[3], Yuhan Zhang[2], Robert H. Lyles[4], Weeberb J. Requia[5], Sindana D. Ilango[6], Howard H. Chang[1,4], Thomas Wingo[7], Rodney J. Weber[3], Joel Schwartz[8]

[#] LS and KS contributed equally.

[1] Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA

[2] Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA

[3] School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

[4] Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

[5] School of Public Policy and Government, Fundação Getúlio Vargas, Brasília, Distrito Federal, Brazil

[6] Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA

[7] Department of Neurology and Human Genetics, School of Medicine, Emory University, Atlanta, Georgia, USA

[8] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

Correspondence to: Liuhua Shi, Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

Telephone: (339)221-8486

E-mail: liuhua.shi@emory.edu

Abstract

Epidemiological evidence suggests air pollution exposure may increase risk of Alzheimer's disease and related dementias (ADRD). However, previous U.S. studies have predominantly focused on hospitalizations, which fails to fully capture ADRD. Here we constructed two national population-based cohorts of those aged 65 and above from the Medicare Chronic Conditions Warehouse (2000-2018), including doctor visits, to investigate the impact of long-term exposure to ambient fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), and ozone (O_3) on dementia and AD incidence, respectively. We identified ~2.0 million incident dementia cases (*N*=12,233,371; dementia cohort) and ~0.8 million incident AD cases (*N*=12,456,447; AD cohort). Per interquartile range (IQR) increase in the 5-year average $PM_{2.5}$ (3.2 μg/m^3), NO_2 (11.6 ppb), and warm-season O_3 (5.3 ppb) over the past 5 years prior to diagnosis, the hazard ratios (HRs) were 1.060 (95% confidence interval [CI]: 1.054, 1.066), 1.019 (95% CI: 1.012, 1.026), and 0.990 (95% CI: 0.987, 0.993) for incident dementias, and 1.078 (95% CI: 1.070, 1.086), 1.031 (95% CI: 1.023, 1.039), and 0.982 (95%CI: 0.977, 0.986) for incident AD, respectively, for the three pollutants. For both outcomes there was strong evidence of linearity in concentration-response relationships for $PM_{2.5}$ and NO_2, suggesting the lack of a clear safe threshold for these health-harmful pollutants. Our study suggests that exposures to $PM_{2.5}$ and NO_2, but not O_3, may increase the incidence of dementia and AD. Improving air quality may reduce the burden of ADRD and promote healthy aging.

Main

Dementia is a major public health issue with substantial health and financial burden, affecting more than 47 million people worldwide[1]. Alzheimer's disease (AD) contributes to about two-thirds of dementia cases and is the sixth leading cause of death in the United States[2]. In response to this devastating public health threat, the National Alzheimer's Project Act was signed into law to overcome dementia, and the National Plan was launched with Goal 1 aiming to prevent and effectively treat dementia (delay onset, slow progression) by 2025[3]. As there are no disease-modifying treatments for the most common types of dementia, it is a top research priority to identify modifiable risk factors for dementia that can be intervened on at the population level.

There is growing evidence associating air pollution with neurodegenerative disease. A systematic review by Peters et al. (2019)[4], found 9 longitudinal studies of air pollution and Alzheimer's disease and related dementias (ADRD). Among them, 5 of 6 showed a positive association between increased exposure to $PM_{2.5}$ and dementia or AD; 4 of 4 showed an association between NO_2 and dementia or AD, while 1 of 3 did so for ozone (O_3). Fu and Yung (2020)[5] published a review and meta-analysis of AD and air pollution, and found a 2-fold excess risk of AD for a 10 $\mu g/m^3$ increase of $PM_{2.5}$ among 6 studies, and no increased risk for NO_2 in four studies, nor for O_3 in three studies. There have been several longitudinal studies since these reviews, with the majority finding positive associations between air pollutants and either dementia or AD[6-14]. A few of these studies examine the associations in US populations, and these studies have almost exclusively used hospitalization as a measure of morbidity[6,7,11,13]. The diagnosis of ADRD, however, likely occurs in doctor visits, and ADRD does not generally result in hospitalizations. Thus, hospitalization records may not well represent either disease incidence or prevalence, and likely leads to an underestimation of the number of cases. In addition,

neuropathologic changes are known to occur many years prior to the diagnosis[15], and the relevant time window in which air pollution might increase the risk of dementia or AD is unclear.

To address these knowledge gaps in studying ADRD incidence in the US, here we constructed a national, population-based cohort study from Medicare data to investigate the impact of long-term exposure to $PM_{2.5}$, NO_2, and O_3 on dementia and AD incidence. To better approximate disease incidence, we required a 5-year "clean" period without events of interest and used all Medicare claims nationwide (2000-2018), including Medicare inpatient and outpatient claims, carrier file (primarily doctor visits), skilled nursing facility, and home health-care claims. We ascertained air pollution based on resident ZIP code, averaged over 5 years prior to diagnosis, which was estimated from national spatiotemporal ensemble exposure models.

Results

Study population characteristics. Table 1 provides descriptive information on the dementia cohort and AD cohort. Both cohorts were followed after requiring a 5-year period without events of interest to better capture disease incidence. There were 12.2 and 12.4 million people in the dementia and AD cohorts, respectively (Table 1). Most of the studied subjects (78.5% and 78.1% for dementia and AD, respectively) entered the cohorts between ages 65 and 74. The median follow-up was 7 years in both cohorts. 16.6% developed dementia (~2.0 million cases), and 6.5% developed AD (~0.8 million cases). More than 90% were not eligible for Medicaid, indicating that most were above the poverty level. A majority of the study population had a comorbidity at some point during follow-up.

Table 1. Descriptive statistics for the study population

	Dementia cohort		AD cohort	
Variables	**Number**	**%**	**Number**	**%**
Number of events	2,025,130	16.6	804,668	6.5
Number of total population	12,233,371	100	12,456,447	100
Total person-years	89,035,081	100	93,278,266	100
Median follow-up years	7		7	
Age at entry (years)				
65-74	9,597,788	78.5	9,734,481	78.1
75-114	2,635,583	21.5	2,721,966	21.9
Sex				
Male	5,023,879	41.1	5,107,942	41.0
Female	7,209,492	58.9	7,348,505	59.0
Race				
White	11,023,202	90.1	11,214,287	90.0
Black	649,081	5.3	666,619	5.4
Other	561,088	4.6	575,541	4.6
Medicaid Eligibility				
Dual-Eligible	800,139	6.5	852,499	6.8
Non-dual Eligible	11,433,232	93.5	11,603,948	93.2
Comorbidity				
Diabetes	4,433,314	36.2	4,590,000	36.8
Hypertension	10,273,506	84.0	10,502,180	84.3
Stroke	1,991,730	16.3	2,137,239	17.2
Heart failure	3,388,540	27.7	3,598,028	28.9
No comorbidities[a]	1,642,674	13.4	1,865,751	15.0
Air pollutants[b]				
Annual $PM_{2.5}$ ($\mu g/m^3$)	9.3 (3.2)		9.3 (3.2)	
Annual NO_2 (ppb)	17.1 (11.6)		17.1 (11.6)	
Warm-season O_3 (ppb)	42.6 (5.3)		42.6 (5.3)	

Note: [a] means none of the above comorbidities; [b] presented as mean concentration (interquartile range).

Air pollution levels. The average annual level of $PM_{2.5}$ of cohort participants during the study period, 9.3 $\mu g/m^3$, was below the US EPA standard of 12 $\mu g/m^3$; NO_2 levels were considerably below the EPA annual standard of NO_2 of 53 ppb. The annual warm-season average O_3 was 42.6 ppb. EPA does not have a standard for annual warm-season O_3. As a reference, the EPA standard for daily maximum of 8-hour average O_3 is 75 ppb. We examined warm-season O_3, because O_3 is more readily formed in the warm season[16], and this metric is often used in long-term epidemiological studies[17]. Figure 1 shows the distribution of the three pollutants across the

US during our study period, as estimated by the exposure models used in our analysis. $PM_{2.5}$ is highest in the eastern US, O_3 in the West, and NO_2 (largely produced by traffic) in urban centers. Further detail on exposure levels can be found in Supplementary Table S1. The three pollutants in our data were only modestly correlated. The Pearson correlations between pollutants (average exposure within the past 5 years) were as follows: $PM_{2.5}$ and O_3 0.22, NO_2 and O_3 0.19, and NO_2 and $PM_{2.5}$ 0.39.

Health effect estimates. Figure 2 provides the main study results from the Cox proportional hazards models stratified by individual characteristics, adjusting for neighborhood-level socioeconomic status (SES), behavioral risk factors, health care capacity variables, and residual temporal and spatial trends (see Methods). An interquartile range (IQR) increase in the 5-year average of the annual $PM_{2.5}$ (3.2 $\mu g/m^3$) in the 5 years prior to diagnosis was associated with an increased risk of dementia (HR=1.061, 95%CI: 1.056, 1.067) in single pollutant models, which changes little in models with other pollutants. An IQR increase in 5-year average NO_2 (11.6 ppb) is associated with an HR of 1.035 (95%CI: 1.028, 1.042) in single pollutant models, dropping to 1.019 (95%CI: 1.012, 1.026) in multi-pollutant models. An IQR increase in the 5-year average of warm-season O_3 (5.3 ppb) has little effect on dementia rates, with HRs of 1.002 (95% CI: 0.998, 1.005) in single pollutant models and 0.990 (95% CI: 0.987, 0.993) in multi-pollutant models.

The findings for AD have a similar pattern to those for dementia, but the hazard ratios are higher per IQR increase, being 1.078 (95%CI: 1.071, 1.086) for $PM_{2.5}$, 1.050 (95%CI: 1.042, 1.059) for NO_2, and 0.999 (95%CI: 0.995, 1.003) for O_3 assessing each pollutant individually. After adjusting for co-pollutants, the effect estimates were similar for $PM_{2.5}$ (HR=1.078, 95%CI: 1.070, 1.086) and attenuated for NO_2 (HR=1.031, 95%CI: 1.023, 1.039), while O_3 is slightly protective (HR=0.982, 95%CI: 0.977, 0.987).

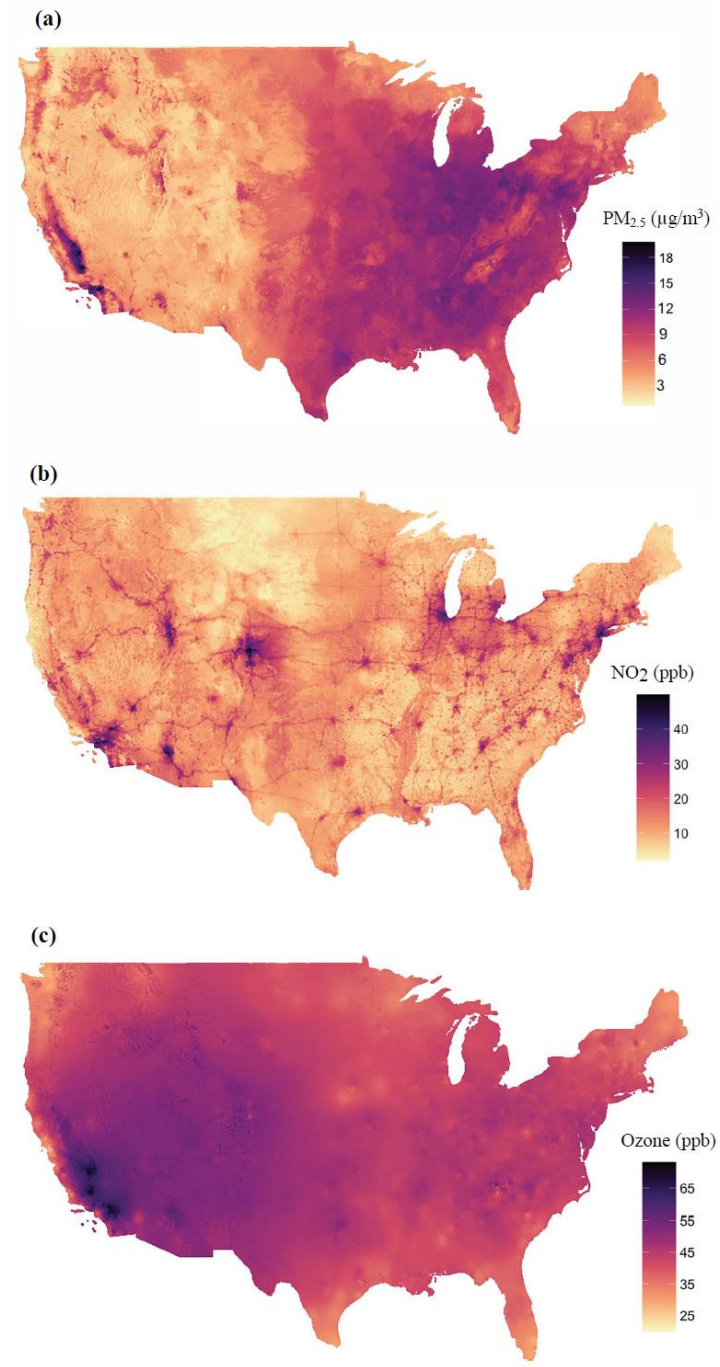

Fig.1| Average concentrations of (a) annual $PM_{2.5}$ (µg/m³), (b) annual NO_2 (ppb), and (c) warm-season O_3 (ppb) across the contiguous United States over the study period.

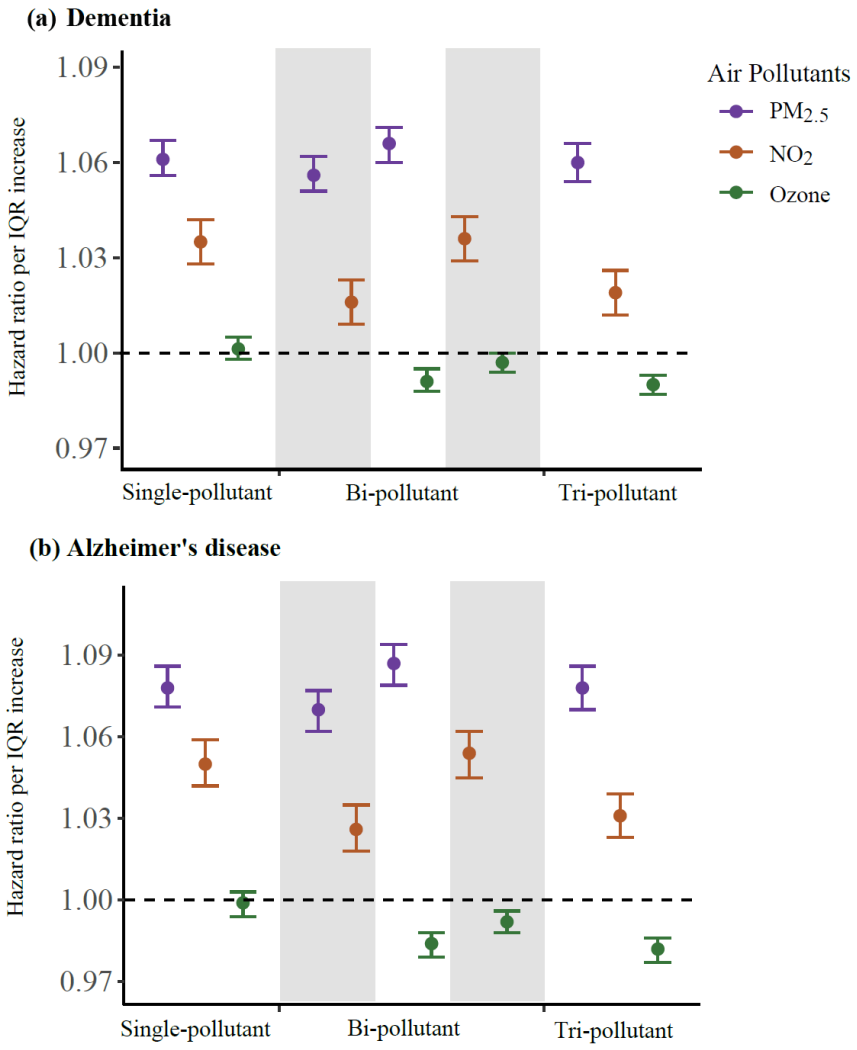

Fig.2| Hazard ratios of dementia or Alzheimer's disease (AD) associated with per IQR increase in annual $PM_{2.5}$ (a), or annual NO_2 (b), or warm-season O_3 (c) concentration. The estimated hazard ratios were obtained using single pollutant. bi-pollutant, and tri-pollutant models.

Concentration-response relationships. Figure 3 presents penalized spline curves for the three pollutants, derived from the tri-pollutant models. The concentration-response (C-R) relationships for $PM_{2.5}$ are approximately linear for both dementia and AD across the exposure distribution, although for AD there is a suggestion of a steeper slope below 8 μg/m^3. For NO_2, the C-R curves for dementia and AD are linear for low concentrations (<25 ppb), and then level off for higher concentrations. The curves for O_3 are essentially flat for both endpoints until high, and rarely occurring concentrations. These results suggest that the adverse effects of $PM_{2.5}$ and

NO_2 on dementia or AD are at least retained, if not strengthened, at low levels of air pollution exposure (e.g., below the WHO air quality guidelines: $PM_{2.5}$ ≤ 10 μg/m^3, NO_2 ≤ 20 ppb). Across the 0.5th to 99.5th percentile of the exposure distribution, $PM_{2.5}$ shows the strongest effect on dementia or AD among all pollutants.

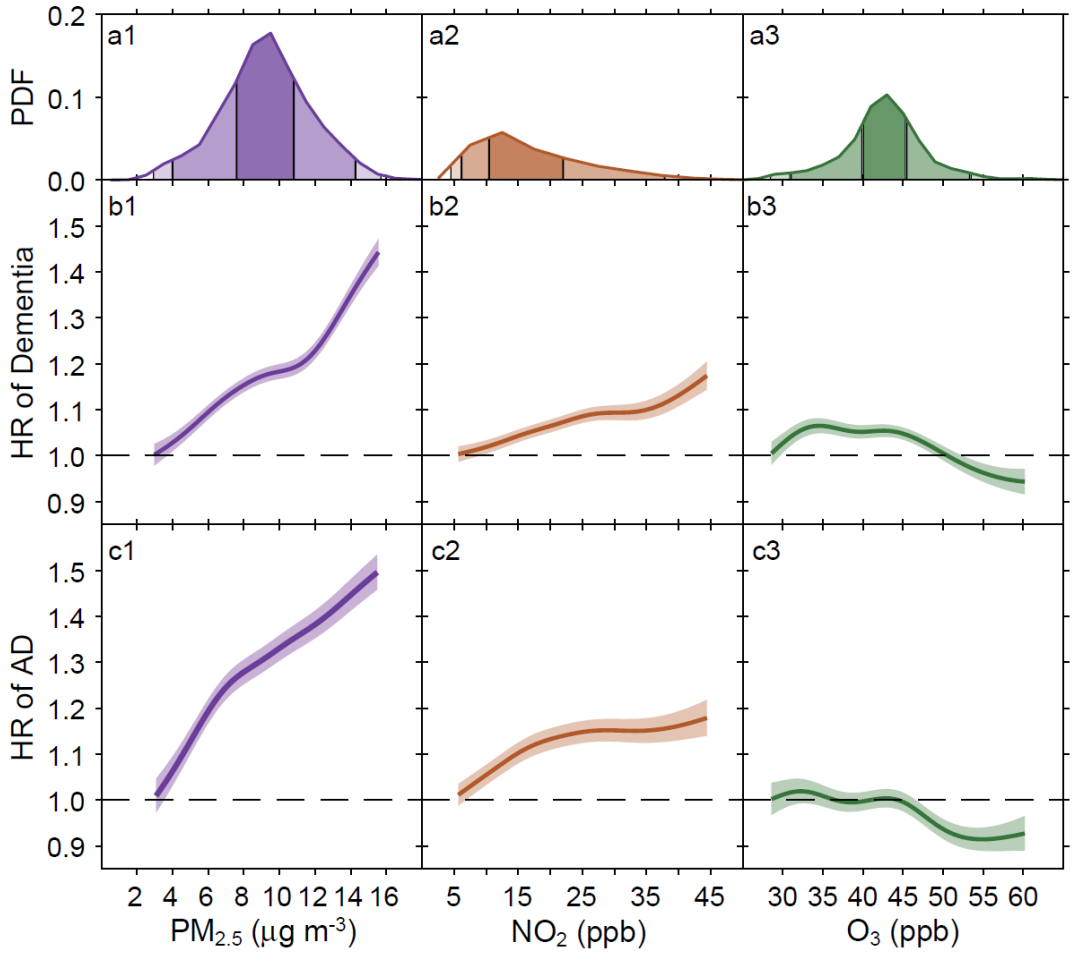

Fig.3| (a) Probability distribution functions (PDF) of long-term $PM_{2.5}$, NO_2, and O_3 exposures; concentration–response relationships between each pollutant and (b) dementia and (c) Alzheimer's disease (AD). The concentration-response curves, derived from the tri-pollutant models, are shown for the concentration ranges between 0.5th to 99.5th percentiles of the pollutants, i.e. with 1% poorly constrained extreme values excluded. Shading areas (from the darkest to the lightest) in panels (a) represent pollutant concentration ranges of the IQR (i.e., 25th to 75th percentiles), 95% (2.5th to 97.5th), and 99% (0.5th to 99.5th), respectively.

Effect modifications. We examined 5 potential effect modifiers (gender, race (white/black/other), Medicaid eligibility, population density (quartiles), and age (<75, ≥75). Figure 4 shows hazard ratios in each subgroup, based on the interaction term between $PM_{2.5}$ or NO_2, and the potential effect modifier. Most marked results were seen for an increased hazard of dementia and AD for blacks vs. whites in relation to both $PM_{2.5}$ and NO_2; a similar pattern was found for those eligible for Medicaid. At the same time, those living in the rural areas (i.e. lowest quartile of population density) were found to have a notably lower association between both dementia and AD and both $PM_{2.5}$ and NO_2. All three of these effect modifiers may reflect a general pattern of increased susceptibility to the effects of $PM_{2.5}$ and NO_2 among those of lower SES and lower education (similar effect modification was seen in Ailshire et al. 2017[18] and Ailshire et al. 2021[19], but not seen in Mortamais et al. 2021[12] or in Cerza et al. 2019[20]). Regarding age, those less than 75 had a markedly stronger association between dementia and both $PM_{2.5}$ and NO_2, while the association was stronger between AD and NO_2 among those older than 75. Finally, we found little evidence of an interaction between pollution and gender in relation to dementia or AD. Figure S1 shows the results for O_3, with relatively few hints of effect modification and all subgroup-specific estimated hazard ratios below one.

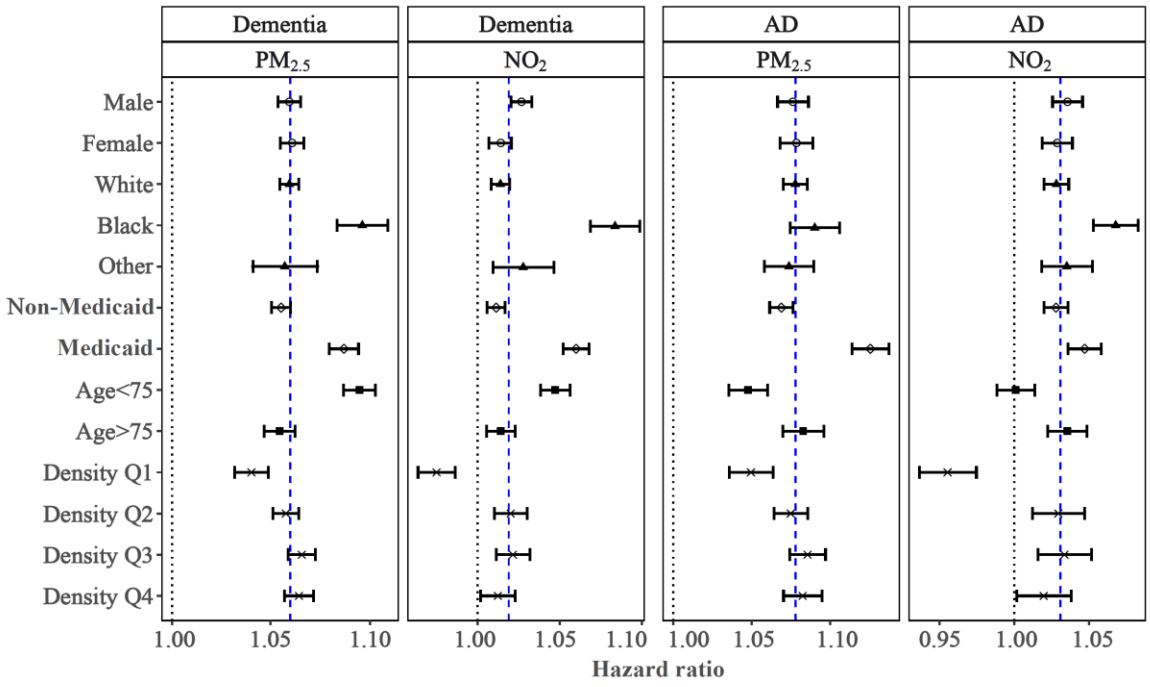

Fig.4| Effect modifications by sex, race, Medicaid eligibility, age, and population density. Results represent the hazard ratios of dementia or Alzheimer's disease (AD), from the tri-pollutant models, per IQR increase in 5-year average $PM_{2.5}$ or NO_2. The blue dashed lines indicate the overall effect estimates for all groups. "Density Q1-Q4" denotes quartiles of population density, i.e., low population density, low-medium population density, medium-high population density, and high population density.

Attributable fraction. The strongest relationship we found with both endpoints was for $PM_{2.5}$ among the three pollutants. If the US $PM_{2.5}$ levels could be lowered by 3.2 $\mu g/m^3$, which is the IQR, then the attributable fraction of dementia and AD due to current exposure levels, based on our main results from tri-pollutant models assuming a linear relationship, would be 6% and 7% respectively. Namely, an estimated 6% of dementia cases and 7% of AD cases would be avoided if $PM_{2.5}$ levels decreased by 3.2 $\mu g/m^3$, which is approximately the difference between our large cities like New York and Chicago and smaller cities like Portland, Buffalo, or Baltimore[21].

Sensitivity analysis. Associations between long-term exposure to $PM_{2.5}$, NO_2, O_3 and dementia or AD were robust to a series of sensitivity analyses. First, a more strict "clean period" by excluding anyone who had a diagnosis for dementia or AD in their first 10 years of follow-up

yielded results similar to the main analyses (Supplementary Table S2). Second, based on this new subcohort (with 10-year clean period), the use of alternative exposure windows (annual exposure 10, 5, 1, or 0 years prior to disease diagnosis, i.e., lags 10, 5, 1, or 0) all support a positive association with $PM_{2.5}$ and NO_2, but not O_3, though HRs varied in magnitude (Supplementary Table S2). For both outcomes, associations with $PM_{2.5}$ and NO_2 were attenuated with increasing lag periods. Third, the observed associations with dementia or AD were not mediated by nor modified by comorbidities, such as diabetes, hypertension, stroke, and heart failure (Supplementary Table S3). Lastly, we assessed the effect of possible outcome misclassification in two ways, one via using a linear regression model based on rates, and the other based on prior estimates of Medicare sensitivity and specificity and estimating the true number of cases within strata. Both methods support the findings from our main analysis i.e., long-term exposure to $PM_{2.5}$ and NO_2, but not O_3, were significantly associated with an increased incidence of dementia and AD; both also suggest that misclassification has somewhat biased our findings to the null (Supplementary Tables S4 and S5).

Discussion

We found elevated hazard ratios for both dementia and AD in relation to $PM_{2.5}$, and less markedly to NO_2, while hazard ratios for warm-season O_3 were not elevated. We did this study in a large US cohort (12 million), with national coverage, and including non-urban areas. For both $PM_{2.5}$ and NO_2, we found a larger effect on AD compared to dementia, which may reflect that fact that dementia includes a wide range of diseases with distinct etiologies, some of which may be unrelated to air pollution, while AD is a subset of dementia and a single disease, for which we found a stronger association. We also found that shorter time windows between exposure ($PM_{2.5}$ or NO_2) and disease showed higher effect estimates, and we posit that it implies an acceleration of an existing process (i.e., accelerating cognitive decline which was already well developed). Moreover, our diagnosis free period requirement provides reasonable

assurance that we are looking at incidence, and use of physician's visits, nursing home data, etc. to ascertain diagnosis avoids missing large numbers of cases, possibly not missing at random, which likely occurs in studies using diagnoses based on hospital admission records.

Some of our models showed a protective effect of O_3. However, when we compare results in Figure 2, we see that in single pollutant models the effect estimate for O_3 was null, while in bi-pollutant models with either $PM_{2.5}$ or NO_2, the effect size for O_3 was pushed below the null (albeit not significantly) and only in the tri-pollutant model was it protective at the conventional 0.05 level. Moreover, in the bi-pollutant models with O_3, the effect sizes for $PM_{2.5}$ and NO_2 increased from their level in the single pollutant models. We interpret this as evidence that there is no effect of O_3, and the protective effect seen in the tri-pollutant model may be due to collinearity.

Our results are broadly consistent with developing literature, which shows relatively consistent effects for $PM_{2.5}$ and NO_2, but less consistent for O_3. We observed an HR of 1.06 for dementia and an HR of 1.08 for AD per 3.2 $\mu g/m^3$ increase in annual $PM_{2.5}$ in single-pollutant models, i.e., equivalent to an HR of 1.10 and an HR of 1.13 per 5 $\mu g/m^3$ increase in $PM_{2.5}$. These values can be compared with our previous Medicare cohort study using hospitalizations[7], reporting an HR of 1.06 for dementia and an HR of 1.17 for AD per 5 $\mu g/m^3$ increase in annual $PM_{2.5}$. A cohort study conducted in Ontario, Canada by Chen et al. (2017)[22] simultaneously accessed the effects of $PM_{2.5}$, NO_2, and O_3 on dementia risks, and they also found significant associations with $PM_{2.5}$ and NO_2, but not O_3. Recent 2018 and 2020 Lancet Commission overviews of modifiable environmental agents associated with disease noted a possible association between air pollutants and dementia, but noted the evidence still preliminary[1,23].

The epidemiologic findings are supported by brain imaging and toxicologic studies. Regarding brain imaging, Shaffer et al. (2021)[24] have found associations between $PM_{2.5}$ and AD

neuropathology upon autopsy, while Laccarino et al. (2020)[25] found an association between $PM_{2.5}$ and positive positron emission tomography (PET) scans for amyloid. Younan et al. (2020)[26] followed 1000 women and found increased cognitive decline on immediate memory/new learning, and increased MRI-determined risk for future AD using a neuroanatomical risk score. These recent findings support earlier neuroanatomical associations found by others[27,28]. Toxicological studies support several plausible biological mechanisms. $PM_{2.5}$ has been consistently linked to oxidative stress, neuroinflammation, systemic inflammation, and all of which, in turn, have been reported as key pathways to AD pathogenesis[27,29,30]. Magnetite nanoparticles from combustion processes have been discovered in the human brain, indicating that particles from urban air pollution can reach the blood-brain barrier (e.g. through interacting with dysfunctional cell)[31].

Our data suggest that lowering air pollution would have a meaningful reduction on AD and dementia that, when applied to the US population, would be an important tool in the fight against dementia and AD. Assuming these associations are causal, our findings suggest that about 6% of dementia cases and 7% of AD cases would be avoided if $PM_{2.5}$ levels decreased by 3.2 $\mu g/m^3$. It should be noted that – assuming our findings are generalizable to other parts of the world – the potential decrease in the burden of AD with lowered air pollution could be greater, considering that the average annual $PM_{2.5}$ level worldwide in 2015 was estimated at 42 $\mu g/m^3$.[32]

Our study has several strengths. To our knowledge, this is the first nationwide, population-based cohort study that focuses on the simultaneous health effects of $PM_{2.5}$, NO_2, and O_3 on dementia and AD. The large sample size gives us ample power to detect effects even though they are small, which is often the case in environmental studies. Second, the use of Medicare claims data that include doctor's visits rather than restricting the data to hospitalizations is likely to include many more cases, given that many cases are never hospitalized, and also cases which are diagnosed earlier and hence better reflect incidence.

Evidence can be found by comparing recent data in another paper about dementia and AD hospitalization in Medicare data,[7] to the data in the current paper. To allow for a fair comparison, we used the same inclusion/exclusion criteria and restricted to the same time period (2000-2016) and geographic region (i.e. the lower 48 states), and we found that using just hospitalization missed nearly 90% of dementia cases and 60% of AD cases, compared to using our current data including doctor's visits (Supplementary Table S6). Third, we used a conservative method by requiring a 5-year "clean" period and restricting analysis to subjects with continuous enrollment in Medicare FFS, and Part A (hospital insurance) and Part B (medical insurance) programs throughout the study period, which can ensure that cases were newly diagnosed and thus better approximate incidence. Lastly, we were able to control for a large number of individual- and neighborhood-level covariates. Inclusion of comorbidities had negligible effect on our results, suggesting that they are unlikely mediators in our studied associations. However, a formal mediation analysis would be important to confirm these findings.

Despite these advantages, some key limitations should be noted. One limitation, typical of using administrative records to identify disease, is potential misclassification of outcome. AD cases in our database represented only about 40% of the dementia cases, suggesting important under-ascertainment of AD, given that AD represents approximately 60-80% of dementia cases[2]. This percentage is quite similar to the findings of Goodman et al. (2017)[33], who found that AD represented 44% of all dementia diagnoses in Medicare data in 2013, including both hospitalizations and doctor visits. It is likely that a large number of our dementia cases, who show no AD diagnosis in Medicare, actually had AD, but physicians did not feel confident to make the more specific diagnosis. This is supported by the findings of Taylor et al. (2009), who compared Medicare data to clinical diagnoses considered as the gold standard, and found that the sensitivity of dementia was 0.85 but was considerably lower, 0.65, for AD.

We have assumed that outcome misclassification is non-differential (conditionally independent of exposure to air pollutants, conditional on confounders); there are no data indicating otherwise. We have conducted two types of sensitivity analyses to adjust for such misclassification of classifying dementia or AD cases as without dementia or AD (false negative, or 1-sensitivity), and the misclassification of non-dementia, non-AD subjects to one of the diseases (false positive, or 1-specificity). Both these methods of adjustment for false negative and false positives were in agreement that our results were likely to under-estimate the true hazard ratios for $PM_{2.5}$ and NO_2 for both dementia and AD.

Another limitation of our study is the potential for exposure error, although the exposure prediction model we used has excellent predictive accuracy[34-36]. Using larger scale ambient air pollutions assigned to individuals has been shown to have a net bias towards the null, consistent with non-differential measurement error, which reflects to some degree classical type of error[37-39]. In addition, our study is subject to unmeasured and residual confounding. While we were able to control for a number of potential confounders at the neighborhood level, we had no individual-level data on SES and education, a limitation implying some mismeasurement of confounders, which may have biased our results (moderately, given that these unmeasured confounders are not likely to act as very strong risk factors for dementia), in an unknown direction. Furthermore, we only studied the Medicare FFS population who enrolled in both Part A and Part B programs, precluding generalizability to the entire US elderly population.

Implications for future research

Our study provides clear evidence that long-term exposure to $PM_{2.5}$ mass is significantly associated with increased ADRD incidence and lowering air pollution potentially has an important public health effect. Future studies of air pollution and dementia in low-to-middle-income countries (LMIC) of which there are few, will be important. Understanding the potential

bias and unmeasured confounding, given the limitations of observational studies, is encouraged. Examining the role of specific pollutant components in ADRD may also be important, because different components of $PM_{2.5}$ (e.g., metals, elemental carbon, organic carbon, sulfate, and nitrate) and different sources of $PM_{2.5}$ (e.g., traffic, industrial, cooking, and biomass burning) may have different neurotoxicities. Better understanding of component-specific and source-specific effects of $PM_{2.5}$ on ADRD could inform pollution control policies on specific sources.

Methods

Study Population

Data were drawn from the Medicare denominator file and the Medicare Chronic Conditions Warehouse (CCW), both from the Centers for Medicare and Medicaid Services (CMS). The denominator file contains enrollment records for each Medicare beneficiary in each year, including age, sex, race, Medicaid eligibility (a proxy for socioeconomic status - SES), the date of death, and ZIP code of residence. CCW provides the date of first occurrence with a dementia or AD diagnosis code across the available Medicare claims. Based on these two Medicare databases, we constructed an open cohort including all Medicare beneficiaries aged 65 and over who were always enrolled (1) in Medicare Fee-for-Service (FFS) program; and (2) in both Medicare Part A (hospital insurance) and Part B (medical insurance) in the contiguous United States between 2000 and 2018. These criteria excluded those with any time in Medicare Advantage (HMO) over the study period since claim records are not available for these beneficiaries and excluded those only enrolled in Medicare Part A or Part B. If we relaxed these restrictions to broaden the cohort, the chance of missing dementia or AD cases among those additional people brought into the analysis would be high.

We created separate datasets for dementia and AD. For each cohort, we further required a “clean” period of 5 years after enrollment, during which there were no diagnosis codes for dementia or AD. By removing potentially prevalent cases in their first five years of follow-up, a diagnosis after that “clean” period more

likely approximates “incidence”. We considered that 5 years was a reasonable period to ensure that a person truly was not demented prior to the Medicare diagnosis; however, we also explored a 10-year clean period in sensitivity analyses. Therefore, study subjects entered the cohort on January 1st of the year following the “clean” period and were followed until first diagnosis of the outcome of interest across all Medicare claims, death, or end of follow-up. We excluded this 5-year clean period from follow-up time to avoid immortal time bias. This study was approved by the Institutional Review Board of Emory University and a waiver of informed consent was granted.

Outcome classification

The primary outcomes of interest for this study were time to 1) all-cause dementia and 2) AD subtype. CCW includes pre-defined indicators for dementia and AD, which are identified using an algorithm that incorporates information from all available Medicare claims (such as inpatient and outpatient claims, Carrier file, skilled nursing facility, and home health-care claims) indicating that an individual was diagnosed with dementia or AD (Chronic condition algorithms 2015[40]). CCW provides the date of first occurrence with a dementia or AD diagnosis code. In the dementia cohort, the outcome dementia was defined as the first occurrence of a diagnosis code of dementia, while for the AD cohort, AD was defined as either 1) the first occurrence of a diagnosis code of AD with no prior diagnosis of dementia, or the first occurrence of a diagnosis code of dementia when there was a subsequent diagnosis code of AD (under the supposition that the original dementia diagnosis was probably AD, given the subsequent AD diagnosis).

Exposure Assessment

High-resolution ambient $PM_{2.5}$, NO_2, and O_3 concentrations at 1-km spatial resolution for the entire United States were derived using spatiotemporal ensemble models that integrated three different machine learning algorithms, including neural networks, random forest, and gradient boosting. The ensemble-based model was calibrated using hundreds of predictors, including satellite measurements, chemical transport model simulations, land-use terms, meteorological variables, and monitoring measurements from the Environmental Protection Agency (EPA) Air Quality Systems (AQS). This ensemble learning

approach yielded strong prediction model performance for each pollutant, with an average cross-validated coefficient of determination (R^2) of 0.89, 0.84 and 0.86 for annual mean $PM_{2.5}$, annual mean NO_2, and warm-season mean maximum 8-hour O_3, respectively[34-36]. We averaged these 1-km resolution predictions for each pollutant at the ZIP code scale across each year, because ZIP Code is the smallest level of geography in the Medicare data. We used the annual averages in each ZIP code, for each calendar year, as the exposure estimates for each Medicare beneficiary according to the ZIP code of residence. All dementia and AD events were linked to exposures averaged over 5 years prior to diagnosis, and any annual residential mobility changes by ZIP code were taken into account, based on their yearly residence in the Medicare database.

Covariates

Individual-level age at entry, sex, race, and Medicaid eligibility were obtained from the Medicare denominator file. We also obtained neighborhood-level covariates in our study. These included ZIP code-level SES variables (population density, % Black population, education, median household income, % owner-occupied housing units, and % population above 65 years of age living below the poverty line), county-level behavioral risk factors (smoking rate and body mass index) and health care capacity variables (number of hospitals and active medical doctors), as well as a geographical region. Data were also available for co-morbidities (diabetes, heart failure, stroke, hypertension) in CCW. These covariates have been associated previously with ADRD and may be associated with air pollution, and hence were candidate confounders to be included in models[41,42].

Statistical Analysis

We fit a series of stratified Cox proportional-hazards models with a generalized estimating equation (GEE)[43] to estimate the associations between long-term exposure to $PM_{2.5}$, NO_2, and O_3 on dementia or AD among the elderly, where the coefficient for the exposure variable was the parameter of interest, and years of follow-up was the time scale. Specifically, we fit single-pollutant, bi-pollutant, and tri-pollutant models and estimated hazard ratios (HRs) per interquartile-range (IQR) increase in the 5-year average of the annual $PM_{2.5}$, NO_2, and warm-season O_3 concentrations in the 5 years prior to diagnosis. All three

pollutants are of interest because some prior literature has shown associations between each of them and dementia[7,20,22,44]. To allow for flexible strata-specific baseline hazard functions, we stratified all models on four individual characteristics, including sex, race (white, black, other), Medicaid eligibility, and 1-year categories of age at study entry. To adjust for potential confounding, we included neighborhood-level SES, behavioral risk factors, and health care capacity variables in our analyses. Potential residual temporal and spatial trends were controlled by respectively including a linear term for calendar years and indicator variables for the geographical region[7].

To assess the shape of the concentration-response (C-R) relationship between each air pollutant and dementia or AD, we respectively fit penalized splines[45] for $PM_{2.5}$, NO_2, and O_3, adjusting for all covariates included in the tri-pollutant models. To identify subpopulations who might be more vulnerable than others, we assessed potential effect modification by sex, race, Medicaid eligibility, age groups (aged 75+ vs. below 75), and urbanicity (quartiles of population density) on the multiplicative scale by including interaction terms between these potential modifiers and pollutants.

Additionally, we estimated the attributable fraction (AF) of dementia and AD cases due to $PM_{2.5}$ air pollution, for those in the US exposed to an additional IQR of $PM_{2.5}$ (a difference of 3.2 $\mu g/m^3$), beyond current levels in US cities with relatively low exposure (i.e., 7 $\mu g/m^3$, the counterfactual), using results from the multi-pollutant model, and using standard AF calculations when the entire population is exposed (RR-1)/RR (see Steenland and Armstrong 2006[46]).

We conducted a series of sensitivity analyses to test the robustness of our main findings. First, we repeated the analyses using a “clean” period of 10 years, i.e., thinking that excluding cases with a diagnosis during their first 10 years of enrollment would increase the probability that we are capturing the first diagnosis and thus more closely estimating disease incidence, albeit at the cost of a smaller number of years of follow-up and cases. Second, using this new subcohort, we assessed alternative exposure time windows by comparing the results using different lags (0-, 1-, 5- and 10-year lags), in which exposure was assigned either as the annual exposure at 10 years prior to case (or the risk set for given cases), or 5 years prior, or 1 or 0 year prior. We posit that if a shorter lag between exposure and disease fits the data best, this would imply an acceleration of an existing process by air pollution, while a longer

lag might indicate the air pollution has an effect in more initial stages of neurodegeneration. Additionally, to evaluate whether the associations we observe can be attributed to comorbidities also linked to air pollution, we additionally adjusted for the comorbidities (including diabetes, hypertension, stroke, and heart failure), and also restricted analyses to subjects without the comorbidities. Finally, we conducted analyses to estimate the effect of possible outcome misclassification in two ways. First, we fit linear regression models for the rate of dementia or AD (events/person-time) with a GEE, which in theory should target an approximately unbiased estimate of the additive effect[47]. Second, we considered the possible effect of outcome misclassification following methods similar to those described by Fox et al. (2005)[48]. We obtained estimates of misclassification parameters from Taylor et al. (2009)[49] and adjusted the observed outcomes for each stratum to match up with the expected true values given pre-specified values for sensitivity and specificity for the outcome classification *(details provided in Supplemental Material)*.

All computational analyses were run on the Rollins High-Performance Computing (HPC) Cluster at Emory University. R software, version 4.0.2, was used for all analyses. A two-sided $P<0.05$ was considered statistically significant.

Data availability

Datasets except the Medicare data reported in the current Article are available on request by qualified scientists. The Medicare data (security level 3 data) are not publicly available, due to restrictions of ethical approval requirements for this study and the Health Insurance Portability and Accountability Act (HIPAA) security rule for security.

Code availability

Custom code that supports the findings of this study is available from the corresponding author on request.

Acknowledgements

This study was supported by the HERCULES Center P30 ES019776 and the Goizueta Alzheimer's Disease Research Center (ADRC) of Emory University (P50 AG025688). The authors acknowledge Dr. Joel Schwartz's lab for providing us with access to their estimated air pollution data and acknowledge Jingxuan Zhao for access to the health care capacity data.

Author contributions

L.S. and K.S. designed research and directed its implementation; L.S., H.L., K.S., and R.H.L. analyzed data, L.S., P.L., Y.Z., and H.L. made the figures and tables, L.S., P.L., W.J.R., and J.S. prepared datasets; L.S., K.S., P.L., R.H.L., S.I., H.C., T.W., J.S. and R.J.W. interpreted the results. L.S. and K.S. lead the writing of the manuscript, with input from all authors.

Competing interests

The authors declare no competing interests.

References

1 Livingston, G. *et al.* Dementia prevention, intervention, and care. *The Lancet* **390**, 2673-2734 (2017).

2 Heron, M. P. Deaths: leading causes for 2017. (2019).

3 Khachaturian, Z. S., Khachaturian, A. S. & Thies, W. The draft “National Plan” to address Alzheimer's disease-National Alzheimer's Project Act (NAPA). *Alzheimer's & Dementia* **8**, 234-236 (2012).

4 Peters, R. *et al.* Air pollution and dementia: a systematic review. *Journal of Alzheimer's Disease* **70**, S145-S163 (2019).

5 Fu, P. & Yung, K. K. L. Air pollution and Alzheimer’s disease: a systematic review and meta-analysis. *Journal of Alzheimer's Disease*, 1-14 (2020).

6 van Wijngaarden, E. *et al.* Neurodegenerative hospital admissions and long-term exposure to ambient fine particle air pollution. *Annals of Epidemiology* **54**, 79-86. e74 (2021).

7 Shi, L. *et al.* Long-term effects of PM2· 5 on neurological disorders in the American Medicare population: a longitudinal cohort study. *The Lancet Planetary Health* **4**, e557-e565 (2020).

8 Smargiassi, A. *et al.* Exposure to ambient air pollutants and the onset of dementia in Québec, Canada. *Environmental Research* **190**, 109870 (2020).

9 Grande, G., Ljungman, P. L., Eneroth, K., Bellander, T. & Rizzuto, D. Association between cardiovascular disease and long-term exposure to air pollution with the risk of dementia. *JAMA neurology* **77**, 801-809 (2020).

10 Ilango, S. D. *et al.* The role of cardiovascular disease in the relationship between air pollution and incident dementia: a population-based cohort study. *International journal of epidemiology* **49**, 36-44 (2020).

11 Lee, M., Schwartz, J., Wang, Y., Dominici, F. & Zanobetti, A. Long-term effect of fine particulate matter on hospitalization with dementia. *Environmental pollution (Barking, Essex: 1987)* **254**, 112926 (2019).

12 Mortamais, M. *et al.* Long-term exposure to ambient air pollution and risk of dementia: Results of the prospective Three-City Study. *Environment International* **148**, 106376 (2021).

13 Nunez, Y. *et al.* Fine Particle Exposure and Clinical Aggravation in Neurodegenerative Diseases in New York State. *Environmental health perspectives* **129**, 027003 (2021).

14 Sullivan, K. J. *et al.* Ambient fine particulate matter exposure and incident mild cognitive impairment and dementia. *Journal of the American Geriatrics Society* (2021).

15 Jack Jr, C. R. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology* **9**, 119-128 (2010).

16 Jacob, D. J. *Introduction to atmospheric chemistry*. (Princeton University Press, 1999).

17 Jerrett, M. *et al.* Long-term ozone exposure and mortality. *New England Journal of Medicine* **360**, 1085-1095 (2009).

18 Ailshire, J., Karraker, A. & Clarke, P. Neighborhood social stressors, fine particulate matter air pollution, and cognitive function among older US adults. *Social science & medicine* **172**, 56-63 (2017).

19 Ailshire, J. & Brown, L. L. The importance of air quality policy for older adults and diverse communities. *Public Policy & Aging Report* **31**, 33-37 (2021).

20 Cerza, F. *et al.* Long-term exposure to air pollution and hospitalization for dementia in the Rome longitudinal study. *Environmental Health: A Global Access Science Source* **18** (2019).

21 USEPA. Air Quality - Cities and Counties. (2020).

22 Chen, H. *et al.* Exposure to ambient air pollution and the incidence of dementia: A population-based cohort study. *Environment international* **108**, 271-277 (2017).
23 Landrigan, P. J. *et al.* The Lancet Commission on pollution and health. *The lancet* **391**, 462-512 (2018).
24 Shaffer, R. M. *et al.* Fine Particulate Matter and Markers of Alzheimer's Disease Neuropathology at Autopsy in a Community-Based Cohort. *Journal of Alzheimer's Disease*, 1-13.
25 Iaccarino, L. *et al.* Association between ambient air pollution and amyloid positron emission tomography positivity in older adults with cognitive impairment. *JAMA neurology* **78**, 197-207 (2021).
26 Younan, D. *et al.* Particulate matter and episodic memory decline mediated by early neuroanatomic biomarkers of Alzheimer's disease. *Brain* **143**, 289-302 (2020).
27 Calderón-Garcidueñas, L. *et al.* Long-term air pollution exposure is associated with neuroinflammation, an altered innate immune response, disruption of the blood-brain barrier, ultrafine particulate deposition, and accumulation of amyloid β-42 and α-synuclein in children and young adults. *Toxicologic pathology* **36**, 289-310 (2008).
28 Cacciottolo, M. *et al.* Particulate air pollutants, APOE alleles and their contributions to cognitive impairment in older women and to amyloidogenesis in experimental models. *Translational psychiatry* **7**, e1022-e1022 (2017).
29 Levesque, S., Surace, M. J., McDonald, J. & Block, M. L. Air pollution & the brain: Subchronic diesel exhaust exposure causes neuroinflammation and elevates early markers of neurodegenerative disease. *Journal of neuroinflammation* **8**, 1-10 (2011).
30 Ranft, U., Schikowski, T., Sugiri, D., Krutmann, J. & Krämer, U. Long-term exposure to traffic-related particulate matter impairs cognitive function in the elderly. *Environmental research* **109**, 1004-1011 (2009).
31 Maher, B. A. *et al.* Magnetite pollution nanoparticles in the human brain. *Proceedings of the National Academy of Sciences* **113**, 10797-10801 (2016).
32 Cohen, A. J. *et al.* Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet* **389**, 1907-1918 (2017).
33 Goodman, R. A. *et al.* Prevalence of dementia subtypes in United States Medicare fee-for-service beneficiaries, 2011–2013. *Alzheimer's & dementia* **13**, 28-37 (2017).
34 Di, Q. *et al.* An ensemble-based model of PM2. 5 concentration across the contiguous United States with high spatiotemporal resolution. *Environment international* **130**, 104909 (2019).
35 Di, Q. *et al.* Assessing NO2 Concentration and Model Uncertainty with High Spatiotemporal Resolution across the Contiguous United States Using Ensemble Model Averaging. *Environmental science & technology* **54**, 1372-1384 (2019).
36 Requia, W. J. *et al.* An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. *Environmental Science & Technology* **54**, 11037-11047 (2020).
37 Kioumourtzoglou, M.-A. *et al.* Exposure measurement error in PM 2.5 health effects studies: a pooled analysis of eight personal exposure validation studies. *Environmental Health* **13**, 2 (2014).
38 Wu, X. *et al.* Causal inference in the context of an error prone exposure: air pollution and mortality. *The Annals of Applied Statistics* **13**, 520-547 (2019).
39 Zeger, S. L. *et al.* Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental health perspectives* **108**, 419-426 (2000).
40 <https://www2.ccwdata.org/web/guest/condition-categories> (
41 Hersi, M. *et al.* Risk factors associated with the onset and progression of Alzheimer's disease: A systematic review of the evidence. *Neurotoxicology* **61**, 143-187 (2017).

42 Anstey, K. J., Ee, N., Eramudugolla, R., Jagger, C. & Peters, R. A systematic review of meta-analyses that evaluate risk factors for dementia to evaluate the quantity, quality, and global representativeness of evidence. *Journal of Alzheimer's Disease* **70**, S165-S186 (2019).
43 Zhang, X. Generalized estimating equations for clustered survival data. (2006).
44 Chen, H. *et al.* Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study. *The Lancet* **389**, 718-726 (2017).
45 Meyer, M. C. Constrained penalized splines. *Canadian Journal of Statistics* **40**, 190-206 (2012).
46 Steenland, K. & Armstrong, B. An overview of methods for calculating the burden of disease due to specific risk factors. *Epidemiology*, 512-519 (2006).
47 Hutcheon, J. A., Chiolero, A. & Hanley, J. A. Random measurement error and regression dilution bias. *Bmj* **340** (2010).
48 Fox, M. P., Lash, T. L. & Greenland, S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *International journal of epidemiology* **34**, 1370-1376 (2005).
49 Taylor Jr, D. H., Østbye, T., Langa, K. M., Weir, D. & Plassman, B. L. The accuracy of Medicare claims as an epidemiological tool: the case of dementia revisited. *Journal of Alzheimer's Disease* **17**, 807-815 (2009).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterials.docx