**Supplementary Appendix for**

**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 Li2, Pengfei Liu3, Yuhan Zhang2, Robert H. Lyles4, Weeberb J. Requia5, Sindana D. Ilango6, Howard H. Chang4, Thomas Wingo7, Rodney J. Weber3, Joel Schwartz8

# 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

**Supplementary tables and figures**

**Table S1**.Annual PM2.5 (g/m3), annual NO2 (ppb), and warm-season O3 (ppb) levels (minimum, maximum, mean, and percentiles) over the study period

**Table S2**. Hazard ratios of dementia or AD per IQR increase in air pollutants among the subpopulation with a 10-year clean period, using varied lag periods

**Table S3**. Hazard ratios of dementia or AD per IQR increase in air pollutants from tri-pollutant models, additionally adjusting for comorbidities

**Table S4**. Rate ratio of dementia or Alzheimer’s Disease (AD) per IQR increase in air pollutants, derived from the linear rate models

**Table S5**. Hazard ratio of dementia or Alzheimer’s Disease (AD) per IQR increase in each pollutant, accounting for potential outcome misclassification via adjusting data for assumed sensitivity and specificity of classification (from Taylor et al. 2009).

**Table S6**. Comparing the dementia and AD cohort using Medicare CCW database versus Medicare inpatient claims (2000-2016)

**Figure S1.** 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 warm-season ozone. 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.

**Table S1**.Annual PM2.5 (g/m3), annual NO2 (ppb), and warm-season O3 (ppb) levels (minimum, maximum, mean, and percentiles) over the study period

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pollutants** | **Min** | **1st** | **5th** | **25th** | **50th** | **75th** | **95th** | **99th** | **Max** | **Mean** |
| **PM2.5** | 0.5 | 3.3 | 4.9 | 7.7 | 9.2 | 10.8 | 13.5 | 15.1 | 25.9 | 9.3 |
| **NO2** | 0.5 | 5.2 | 8.0 | 10.5 | 14.8 | 22.0 | 34.0 | 41.8 | 122.2 | 17.1 |
| **O3** | 19.8 | 29.1 | 33.4 | 40.0 | 42.7 | 45.4 | 51.2 | 57.3 | 77.7 | 42.6 |

**Table S2**. Hazard ratios of dementia or AD per IQR increase in air pollutants among the subpopulation with a 10-year clean period, using varied lag periods

|  |  |  |  |
| --- | --- | --- | --- |
|  | **PM2.5** | **NO2** | **O3** |
| **Dementia – Single pollutant** |
| Lag 0  | 1.062 (1.055, 1.069) | 1.029 (1.022, 1.036) | 1.003 (0.999, 1.006) |
| Lag 1 | 1.052 (1.046, 1.059) | 1.021 (1.015, 1.028) | 0.998 (0.994, 1.001) |
| Lag 5  | 1.042 (1.037, 1.046) | 1.018 (1.012, 1.023) | 1.004 (1.001, 1.007) |
| Lag 10  | 1.028 (1.025, 1.032) | 1.011 (1.007, 1.015) | 1.003 (1.000, 1.005) |
| Average (lags 0-5) | 1.053 (1.047, 1.059) | 1.022 (1.015, 1.029) | 0.999 (0.995, 1.002) |
| **AD – Single pollutant** |
| Lag 0  | 1.096 (1.082, 1.111) | 1.076 (1.062, 1.089) | 0.991 (0.984, 0.999) |
| Lag 1 | 1.091 (1.079, 1.103) | 1.064 (1.052, 1.076) | 0.986 (0.979, 0.992) |
| Lag 5  | 1.077 (1.068, 1.087) | 1.052 (1.042, 1.063) | 0.996 (0.991, 1.002) |
| Lag 10  | 1.066 (1.058, 1.073) | 1.042 (1.033, 1.050) | 0.992 (0.987, 0.998) |
| Average (lags 0-5) | 1.079 (1.070, 1.088) | 1.049 (1.039, 1.059) | 0.996 (0.991, 1.001) |
| **Dementia – Multi-pollutant** |
| Lag 0  | 1.060 (1.053, 1.067) | 1.016 (1.009, 1.023) | 0.995 (0.991, 0.998) |
| Lag 1 | 1.053 (1.047, 1.060) | 1.012 (1.005, 1.018) | 0.990 (0.987, 0.994) |
| Lag 5  | 1.041 (1.036, 1.047) | 1.007 (1.001, 1.012) | 0.997 (0.994, 0.999) |
| Lag 10  | 1.030 (1.026, 1.035) | 1.001 (0.996, 1.006) | 0.995 (0.993, 0.998) |
| Average (lags 0-5) | 1.055 (1.048, 1.062) | 1.009 (1.002, 1.016) | 0.991 (0.987, 0.994) |
| **AD – Multi-pollutant** |
| Lag 0  | 1.090 (1.075, 1.104) | 1.063 (1.050, 1.077) | 0.976 (0.968, 0.983) |
| Lag 1 | 1.090 (1.077, 1.103) | 1.053 (1.041, 1.066) | 0.970 (0.964, 0.977) |
| Lag 5  | 1.078 (1.067, 1.088) | 1.038 (1.027, 1.048) | 0.979 (0.974, 0.985) |
| Lag 10  | 1.073 (1.065, 1.081) | 1.022 (1.013, 1.030) | 0.973 (0.968, 0.978) |
| Average (lags 0-5) | 1.078 (1.068, 1.088) | 1.033 (1.023, 1.043) | 0.981 (0.976, 0.986) |

**Table S3**. Hazard ratios of dementia or AD per IQR increase in air pollutants from tri-pollutant models, additionally adjusting for comorbidities

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outcome** | **Models** | **PM2.5** | **NO2** | **O3** |
| **Dementia** | Main model + diabetes | 1.058 (1.052, 1.064) | 1.004 (0.997, 1.010) | 0.993 (0.990, 0.996) |
| Main model + hypertension | 1.051 (1.045, 1.057) | 1.014 (1.008, 1.021) | 0.992 (0.988, 0.995) |
| Main model + stroke | 1.051 (1.046, 1.057) | 1.013 (1.007, 1.020) | 0.991 (0.988, 0.994) |
| Main model + heart failure | 1.052 (1.046, 1.058) | 1.003 (0.997, 1.009) | 0.992 (0.988, 0.995) |
| Main model among no comorbidities | 1.061 (1.048, 1.075) | 0.922 (0.910, 0.935) | 1.029 (1.021, 1.037) |
| **AD** | Main model + diabetes | 1.077 (1.069, 1.084) | 1.022 (1.014, 1.030) | 0.983 (0.979, 0.988) |
| Main model + hypertension | 1.072 (1.064, 1.080) | 1.027 (1.019, 1.036) | 0.983 (0.978, 0.987) |
| Main model + stroke | 1.071 (1.064, 1.079) | 1.027 (1.018, 1.035) | 0.983 (0.978, 0.987) |
| Main model + heart failure | 1.074 (1.066, 1.081) | 1.022 (1.014, 1.031) | 0.983 (0.978, 0.987) |
| Main model among no comorbidities | 1.062 (1.044, 1.080) | 0.900 (0.882, 0.917) | 1.021 (1.011, 1.032) |

**Sensitivity analysis on outcome misclassification.**

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 generalized estimating equation, which resulted in an estimate of the additive effect less sensitive to bias. The advantage is that in a linear model random misclassification of the outcome would tend to absorb into the residual errors of the linear model for the true counts so that the outcome misclassification should be less likely to produce markedly biased coefficient estimates (e.g., Hutcheon et al., 2020)1. In the linear rate model, the outcome is rate, and the coefficient of exposure is the incremental probability of the event (i.e., the increase in rate) for a unit change in exposure.

Second, we considered the possible effect of outcome misclassification, in a manner similar to the methods described by Fox et al. (2005)2. We used estimates of misclassification from Taylor et al. (2009)3 and adjusted the observed case counts for each zip code in the stratified Poisson model to match up with the expected true values given pre-specified values for sensitivity and specificity for the outcome classification.

We first reduced the number of observations in the stratified Poisson model by restricting to stratification by age and race, rather than all the co-variates included in the original Cox model, in order to avoid smaller strata with 0 disease counts. We then switched focus from person-time to case counts, in order to adjust the observed proportion of cases via our correction for misclassification. We also restricted the data so that each person was associated with only one ZIP code, which was the ZIP code where he/she had lived the longest during the follow-up period. Then each person was assigned the exposure level of that ZIP code. We used the log of the number of people in each ZIP code as the offset associated with the corresponding record in the regression analysis. Using this new data, the original Cox model exposure coefficients, as well as the new exposure coefficient using a simplified Poisson model based on people rather than person-time, are shown in Table S5, based on the tri-pollutants models for both dementia and AD. Finally, we corrected the data for misclassification using the observed error-prone case counts of dementia and AD in each ZIP code, and the sensitivity and specificity for disease classification taken from Taylor et al. (2009), who compared Medicare diagnoses to a gold standard based on clinical diagnoses for 794 members of the Aging Demographics and Memory Study (ADAMS). Taylor et al. (2009)3 estimated the sensitivity and specificity of Medicare claims as 0.85 and 0.89 for dementia, and 0.64 and 0.95 for AD. The correction was done using the formula below (Bross 1954)4:

 p̂i = (p̂i\*+ SP - 1)/(SE + SP - 1),

where p̂i\* is the observed proportion of cases in each ZIP code i, SP and SE are specificity and sensitivity taken from Taylor et al. (2009)3, and p̂i is the corrected proportion of cases in each ZIP code i. As the formula above yields a negative number of corrected cases when the proportion of observed cases is less than 1-SP, we restricted this correction for dementia or AD to ZIP codes where the observed proportion was greater than 1-SP (11% for dementia, 5% for AD); for strata with an observed proportion less than 1-SP, the corrected case count was set to 0. Under these conditions, Table S5 gives the corrected values of the exposure coefficients for both dementia and AD, Corrected HR values from the original Cox model were obtained by taking the ratio of the corrected vs uncorrected exposure coefficients from the Poisson model described above (from models 3 and 2 in Table s5), and then multiplying the exposure coefficient from the original Cox model by that ratio, and obtaining the HR for an IQR increase in PM2.5 using that corrected Cox model coefficient.

**Table S4**. Rate ratio of dementia or Alzheimer’s Disease (AD) per IQR increase in air pollutants, derived from the linear rate models.

|  |  |  |
| --- | --- | --- |
|  | **Cox model** | **Linear rate model** |
| **Dementia** |
|  PM2.5 | **1.060** (1.054, 1.066) | **1.067** (1.062, 1.071) |
|  NO2 | **1.019** (1.012, 1.026) | **1.064** (1.058, 1.069) |
|  O3 | **0.990** (0.987, 0.993) | **0.990** (0.987, 0.993) |
| **AD** |
|  PM2.5 | **1.078** (1.070, 1.086) | **1.105** (1.098, 1.113) |
|  NO2 | **1.031** (1.023, 1.039) | **1.071** (1.063, 1.080) |
|  O3 | **0.982** (0.977, 0.986) | **0.978** (0.974, 0.983) |

Note: The baseline incidence rate of dementia or AD was calculated based on the number of events and total person years in Table 1.

**Table S5**. Hazard ratio of dementia or Alzheimer’s Disease (AD) per IQR increase in each pollutant, accounting for potential outcome misclassification via adjusting data for assumed sensitivity and specificity of classification (from Taylor et al. 20093).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Original Cox model****(model 1)** | **Poisson model with number of people instead of person-time\*****(model 2)\*** | **Corrected hazard ratio from Poisson model with number of people instead of person-time****(model 3)\*\*** | **Corrected hazard ratio from original Cox model** **(model 4)\*\*\*** |
| **Dementia** |  |
|  PM2.5 | **1.060** (1.054, 1.066) | 1.063 (1.060, 1.066) | 1.084 (1.077, 1.090) | **1.081** (1.074, 1.086) |
|  NO2 | **1.019** (1.012, 1.026) | 1.030 (1.027, 1.033) | 1.041 (1.033, 1.048) | **1.026** (1.021, 1.030) |
|  O3 | **0.990** (0.987, 0.993) | 0.979 (0.977, 0.980) | 0.968 (0.965, 0.972) | **0.985** (0.983, 0.987) |
| **AD** |  |
|  PM2.5 | **1.078** (1.070, 1.086) | 1.102 (1.097, 1.107) | 1.303 (1.273, 1.344) | **1.226** (1.204, 1.256) |
|  NO2 | **1.031** (1.023, 1.039) | 1.055 (1.050, 1.060) | 1.176 (1.137, 1.209) | **1.097** (1.076, 1.114) |
|  O3 | **0.982** (0.977, 0.986) | 0.963 (0.960, 0.966) | 0.901 (0.891, 0.912) | **0.950** (0.944, 0.956) |

\*model with stratification by age and race, and exposure assigned to ZIP code of longest residence during follow-up period. Confidence intervals are overly narrow for corrected hazard as they do not reflect uncertainty due to the degree of correction.

\*\* same model as model 2 but corrected for misclassification described in text above.

\*\*\* Original coefficient from Cox model, corrected by multiplying the ratio of the exposure coefficients for model 3/model 2 by the original Cox model exposure coefficient, then multiplying that corrected coefficient by the IQR and exponentiating to get the corrected HR. 95% CIs for models 3 and 4 were calculated from 50 bootstrap samples having the same data size with the modeled dataset.

**Table S6**. Comparing the dementia and AD cohort using Medicare CCW database versus Medicare inpatient claims (2000-2016)

|  |  |  |
| --- | --- | --- |
|  | **AD** | **Dementia** |
| Current study using Medicare CCW database (2000-2016) |
|  Number of claims | 5,646,187 | 11,121,272 |
|  Total person-years | 403,149,214 | 379,921,997 |
| Previous study using Medicare inpatient claims5 (2000-2016) |
|  Number of admissions | 2,490,431 | 1,233,132 |
|  Total person-years | 475,820,277 | 478,636,053 |



**Figure S1.** **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 warm-season ozone.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.

References

1 Hutcheon, J. A., Chiolero, A. & Hanley, J. A. Random measurement error and regression dilution bias. *Bmj* **340** (2010).

2 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).

3 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).

4 Bross, I. Misclassification in 2 x 2 tables. *Biometrics* **10**, 478-486 (1954).

5 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).