Environ Health Perspect, DOI: 10.1289/ehp14723, Accepted 3/4/2025

Estimating the effects of hypothetical ambient $PM_{2.5}$ interventions on the risk of dementia using the parametric g-formula in the UK Biobank cohort

Chengyi Lin[1,2], Riyang Liu[1,2], Caroline Sutton[1,2], Andrew Thomas DeWan[3,4], Laura Forastiere[5], Kai Chen[1,2]

[1] Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA

[2] Yale Center on Climate Change and Health, Yale School of Public Health, New Haven, CT, USA

[3] Yale Center for Perinatal, Pediatric and Environmental Epidemiology, Yale School of Public Health, New Haven, CT, USA

[4] Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA

[5] Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

Corresponding author: Kai Chen, PhD, Department of Environmental Health Sciences, Yale School of Public Health, 60 College St, New Haven, CT 06520-8034 (kai.chen@yale.edu).

Conflict of Interest Disclosures: The authors declare they have no conflicts of interest related to this work to disclose.

Abstract

Background: Emerging evidence identifies ambient particulate matter with an aerodynamic diameter ≤ 2.5μm ($PM_{2.5}$) as a modifiable risk factor for dementia, but the potential health benefits gained by enacting regulations that reduce $PM_{2.5}$ remain unclear.

Objectives: Our aim was to estimate the total effects of hypothetical ambient $PM_{2.5}$ interventions starting in late life on the risk of dementia in a cohort using the parametric g-formula.

Methods: We used data from 291,495 participants in the UK Biobank cohort who were free of dementia and aged ≥ 55 years at baseline (2010). We estimated the total effects of hypothetical ambient $PM_{2.5}$ interventions (achieving annual average standards of 12 μg/m^3, 10 μg/m^3, and 9 μg/m^3) from 2010 to 2019 on the risk of dementia by calculating the difference between the estimated 10-year risk of dementia under a specified hypothetical intervention and the risk under no intervention using the parametric g-formula.

Results: Compared with no intervention, the estimated 10-year risk difference of dementia was -0.54 per 1000 population (95%CI: -1.00, -0.10), -1.36 per 1000 population (95%CI: -2.44, -0.25), -1.92 per 1000 population (95%CI: -3.39, -0.33), with $PM_{2.5}$ interventions achieving annual average standards of 12 μg/m^3, 10 μg/m^3, and 9 μg/m^3, respectively.

Discussion: The estimated 10-year risk of dementia decreased if the individual ambient $PM_{2.5}$ exposure was reduced due to more stringent $PM_{2.5}$ standards in late life compared to the natural course without intervention on ambient $PM_{2.5}$ exposure. Our findings, obtained using the parametric g-formula — a causal inference method that can directly evaluate the impact of hypothetical interventions — suggest that policies reducing ambient $PM_{2.5}$ pollution may lower the risk of dementia among UK Biobank participants who would experience more stringent ambient $PM_{2.5}$ standards in late life.

Introduction

Results of epidemiological[1] and toxicological studies suggest that particulate matter with an aerodynamic diameter less than or equal to 2.5μm ($PM_{2.5}$) is a potentially modifiable risk factor for dementia, with evidence of oxidative stress,[2] inflammation,[3] elevated Alzheimer's disease-related biomarkers,[4-7] and neurodegeneration[8] associated with $PM_{2.5}$ exposure.

Ambient $PM_{2.5}$ levels have decreased over time in many regions due to air pollution regulations.[9-11] As policymakers debate implementing policies to further reduce $PM_{2.5}$ concentrations, weighing the cost of $PM_{2.5}$ reduction against the associated health benefits is crucial. However, the effect of ambient $PM_{2.5}$ intervention on dementia risk still requires assessment.

The parametric g-formula is a causal method that can be applied to time-varying exposures and confounders. Compared with the traditional health impact assessment using the exposure-response relationships estimated from Cox regression and simplified assumptions on the uniformed change in exposure levels for the entire study population, the parametric g-formula directly evaluates the effects of hypothetical regulatory intervention on health outcomes and can estimate the absolute risk reduction (risk difference).[12] Although widely used in occupational[13-15] and nutritional[16, 17] epidemiology, it was only recently introduced in environmental epidemiology to quantify the effects of hypothetical ambient air pollution interventions on the risk of mortality.[18, 19]

The UK Biobank is a cohort with a large participant base, multiple sources of health records, and residential address history. Several studies have leveraged this cohort and found associations between long-term $PM_{2.5}$ exposure and dementia.[20-22] However, these studies typically relied on the time-fixed $PM_{2.5}$ exposure, overlooking the time-varying nature of the long-term $PM_{2.5}$ exposure. Additionally, the contrast of cause-specific cumulative incidence provides

a clearer causal interpretation than the cause-specific hazard ratio from Cox model, as it compares outcome distributions between groups that would be exchangeable under different interventions.[23] Thus, we used the parametric g-formula to estimate the total effects of hypothetical ambient $PM_{2.5}$ threshold interventions in late life on the 10-year risk of dementia in the UK Biobank cohort from 2010 to 2019.

Method

Study Population

This study used data from the UK Biobank – a prospective cohort that recruited more than 500,000 participants between 2006 to 2010 in the United Kingdom. The participants, aged 37 to 73 years at recruitment, were from various socioeconomic backgrounds.[24] All participants provided consent to the linkage of their electronic medical health records. The UK Biobank received ethical approval as a Research Tissue Bank from the North West Multi-Centre Research Ethics Committee (21/NW/0157).

To assess the effect of hypothetical ambient $PM_{2.5}$ intervention from 2010 to 2019 on the 10-year risk of dementia, participants were excluded from the study if they were diagnosed with dementia (all-cause), dead, lost to follow-up, or less than 55 years old at the baseline year 2010. We further excluded participants who were missing residential address history for the derivation of environmental exposure and neighborhood-level sociodemographic status for the years 2001 to 2009 (**Figure S1, Table S1**). Participants were followed until the earliest occurrence of diagnosis of all-cause dementia, death, loss to follow-up, or end of follow-up on December 31st, 2019.

Historical Residential Location

The UK Biobank Team derived data on participants' residential location change history, (i.e., the date of each location change and the corresponding coordinates of the latest location) at a 1km grid resolution in the year 2017, using detailed residential address from various sources including General Practitioner (GP) systems, current address provided by participants during UK Biobank assessment center visits, and address changes directly reported to the UK Biobank by participants.[25] We expanded location history data by incorporating the information on "Length of time at current address" (data-field 669) obtained during each assessment center visit, and we assumed that the residential address remained unchanged from the beginning of the year 2001 to the earliest record date for individuals whose earliest record date falls within the year 2001.

$PM_{2.5}$ Assessment

We derived ambient $PM_{2.5}$ exposure (in $\mu g/m^3$) from a newly developed model of monthly average $PM_{2.5}$ concentrations at a 1×1 km^2 spatial resolution.[26] The model used the light gradient-boosting machine (LightGBM) framework to capture reliable long-term $PM_{2.5}$ spatiotemporal associations in the UK, incorporating multi-source geographical predictors. The model shows robust performance with a by-year cross-validation (CV) coefficient of determination (R^2) of 0.82 at the monthly level and a root mean square error (RMSE) of 2.13 $\mu g/m^3$.[26] We obtained individual-level monthly exposure by bilinear interpolation from the values of the four nearest grids of each residential address location (**Figure S2**). We obtained the time-varying annual mean exposure by averaging the monthly exposure values.

Covariates

We obtained year and month of birth, sex (female or male), educational attainment (categorized as with or without a college degree, or unknown, where the category of "unknown" comprises individuals who selected "prefer not to answer" or had missing data), average total household income before tax (categorized as less than 18,000£, 18,000£ to 30,999£, 31,000£ to 51,999£, 52,000£ to 100,000£, greater than 100,000£, don't know, or as unknown, which includes individuals who selected "prefer not to answer" or had missing data) from questionnaire data collected at the first visit to assessment centers. The modeled annual average exposure to nitrogen dioxide and nitrogen oxides for the year 2010 was developed by the European Study of Cohorts for Air Pollution Effects (ESCAPE) using a Land Use Regression model.[27] These modeled exposures to nitrogen dioxide and nitrogen oxides were then linked to the residential addresses of each participant. For the time-varying confounders, we calculated the annual mean ambient temperature in Celsius based on ERA5 monthly mean data at a spatial resolution of 0.25° x 0.25°.[28] The individual-level temperature was extracted from the grid cell containing the residential location. We obtained the annual Gross disposable household income (GDHI) per head as indices at International Territorial Levels 3 (ITL3), from the Office for National Statistics (ONS) 1997 to 2021 edition of the regional gross disposable household income dataset. GDHI per head is an indicator of economic welfare that allows for regional comparisons.[29] GDHI per head index is calculated as 100 times the ratio of an area's GDHI per head to the UK average. Annual population density at ITL3 was calculated as the proportion of mid-year residential population, obtained from ONS, to the corresponding regional area. We also adjusted the pre-baseline value of the aforementioned time-varying confounders measured as the 9-year average from 2001 to 2009. We chose these covariates as the confounders based on substantive knowledge of the relationship between long-term exposure to $PM_{2.5}$ and dementia from the literature.[30-32]

Outcome Assessment

The earliest date of diagnosis with all-cause dementia was identified from primary care records and primary or secondary diagnoses in hospitalization data. While hospitalization data are linked for all participants, primary care data are currently available for approximately 45% of participants. All-cause dementia diagnoses were classified using the Read codes (**Table S2**) and the International Classification of Diseases 9th Revision (ICD-9) codes 290.2, 290.3, 290.4, 294.1, 331.0, 331.1, 331.2, 331.5 in older Scottish hospital records and ICD-10 codes A81.0, I67.3, F00, F01, F02, F03, F05.1, G30, G31.0, G31.1 and G31.8. Since death is a semi-competing event for dementia,[33] we obtained the date of death from linked death registry data. In a validation study of approximately 17,000 Scottish participants, the positive predictive value (PPV) for all-cause dementia was estimated to be 86.8% (95% CI: 78.8%, 92.6%) using primary care data, and 87.3% (95% CI: 75.5%, 94.7%) using hospital admission data.[34]

Hypothetical $PM_{2.5}$ Interventions

We estimated the 10-year risk of dementia under three hypothetical ambient $PM_{2.5}$ interventions assigned to all eligible participants starting in 2010. These hypothetical interventions are threshold interventions, where individual-level ambient $PM_{2.5}$ exposures are reduced to the threshold value if they exceed this limit during a follow-up interval, otherwise, they remain unchanged.[35] The hypothetical interventions were compared with no intervention on $PM_{2.5}$. We considered three thresholds: (1) England's 2028 interim target of an annual average of 12 $\mu g/m^3$;[36] (2) England's 2040 and Scotland's 2020 targets of an annual average of 10 $\mu g/m^3$;[36] and (3) the primary annual standard by U.S. Environmental Protection Agency of an annual average of 9 $\mu g/m^3$.[37]

Statistical Analysis

The total effect of $PM_{2.5}$ intervention was used as the counterfactual estimand in this study with death as the competing event.[38] We estimated the risk of dementia that would have been observed if all participants had been assigned to a specified $PM_{2.5}$ threshold, without eliminating the competing event. The risk difference (RD) between the risk under a specified $PM_{2.5}$ intervention and the risk under no intervention represents the total effect of $PM_{2.5}$ intervention on the risk of dementia through all pathways including the potential indirect pathway mediated by increased survival. This is different from the risk of dementia under elimination of the competing event, which is the risk of dementia under a hypothetical intervention that eliminates death. The corresponding RD represents the direct effect of a specified $PM_{2.5}$ intervention on the risk of dementia not mediated by prolonged survival. Since the occurrence of the competing event (i.e., death) precludes the event of dementia,[23] we also estimated the effects of the same hypothetical $PM_{2.5}$ interventions on the risk of all-cause mortality using the parametric g-formula.

These risks can be estimated by the parametric g-formula under the assumptions of no unmeasured confounding, positivity, and consistency.[39] The details of the parametric g-formula algorithm can be found in a paper by McGrath et al.[39] The outline of the steps are as follows. First, we fitted linear regression models for each time-varying covariates separately(i.e., in the order of annual mean temperature, annual population density, annual GDHI per head index, and annual $PM_{2.5}$ exposure), which were conditional on the time-fixed confounders (i.e., age, sex [female or male], and educational attainment [with or without a college degree, or unknown], the pre-baseline values of mean temperature, population density, GDHI per head index, and $PM_{2.5}$ exposure measured as the 9-year average from 2001 to 2009), the previous-year values of the time-varying covariates (i.e., annual mean temperature, annual population density, annual GDHI per head index, and annual $PM_{2.5}$ exposure), the same-year values of time-varying covariates set to occur before the covariate of interest, and the categorical time interval index. Second, we

fitted one logistic regression model to estimate the hazard of dementia conditional on the time-fixed confounders (i.e., age, sex [female or male], and educational attainment [with or without a college degree, or unknown], the pre-baseline values of mean temperature, population density, GDHI per head index, and $PM_{2.5}$ exposure measured as the 9-year average from 2001 to 2009), the previous-year and same-year values of the time-varying covariates (i.e., annual mean temperature, annual population density, annual GDHI per head index, and annual $PM_{2.5}$ exposure), and the categorical time interval index, and another logistic regression model to estimate the hazard of mortality (the competing event) conditional on the time-fixed confounders (i.e., age, sex [female or male], and educational attainment [with or without a college degree, or unknown], the pre-baseline values of mean temperature, population density, GDHI per head index, and $PM_{2.5}$ exposure measured as the 9-year average from 2001 to 2009), the same-year and previous one to four years' values of the time-varying covariates (i.e., annual mean temperature, annual population density, annual GDHI per head index, and annual $PM_{2.5}$ exposure), the interaction terms between the categorical time interval index and time-varying covariates from previous one to four years, and the categorical time interval index (detailed specification in **Table S3**). Third, we used Monte Carlo simulation with 500,000 generated participants where the values of baseline covariates were resampled with replacement from those of the original eligible participants and the $PM_{2.5}$ value was set to the specified $PM_{2.5}$ threshold. The values of the time-varying covariates at the next time point were simulated iteratively from the models for the time-varying covariates estimated in the first step. The hazard of dementia and the hazard of death at each follow-up time were estimated based on the outcome model and the competing event model, with the respective covariate values simulated in the previous steps. At each time point, the risk of dementia for each participant was calculated as the cumulative sum of the product of the hazard of dementia and the probability of surviving without dementia at all previous time points. The standardized risk of dementia under a specified $PM_{2.5}$ intervention at each follow-up time was obtained by calculating the average

risk of dementia among all participants in the simulation. The 95% confidence intervals used were the 2.5th and 97.5th percentiles of a nonparametric bootstrap distribution generated from 200 bootstrap samples. Similarly, we conducted another parametric g-formula analysis to estimate the risk of mortality under hypothetical $PM_{2.5}$ interventions. The linear regression models for the time-varying covariates and the logistic regression model to estimate the hazard of mortality are the same as those used in the above parametric g-formula analysis with dementia as the outcome. The standardized risk of mortality was obtained by calculating the average of the risk of mortality (the cumulative sum of the product of the hazard of mortality and the probability of surviving at all previous time points) for each participant, in a 500,000-participant Monte Carlo simulation. The parametric g-formula method was implemented using the "gfoRmula" package in R version 4.3.2.[39] Next we conducted subgroup analysis by sex and by age at the baseline year 2010 (≥ 65 years or < 65 years), to estimate modification of the association between $PM_{2.5}$ and dementia by sex and age.[40]

We conducted several sensitivity analyses to test the robustness of our findings. To assess the sensitivity of the estimates to residual confounding, we 1) adjusted for the 10-year average of time-varying covariates from the pre-baseline value of 2000 to 2009 instead of the 9-year average (2001-2009); 2) adjusted for annual cold season (December to February) and warm season (June to August) mean temperature instead of the annual mean temperature; 3) additionally adjusted for the average total household income before tax; and 4) additionally adjusted for the modeled annual average exposure to nitrogen dioxide and nitrogen oxides for the year 2010. To assess the validity of model specifications, we 1) tested a different order of time-varying confounders (i.e., in the order of "annual GDHI per head index", "annual mean temperature", "annual population density" instead of the order of "annual mean temperature", "annual population density", "annual GDHI per head index") in the models fitted for confounders; 2) used $PM_{2.5}$ exposures in a log-transformed scale instead of the original scale; and 3)

compared observed and parametric g-formula estimated values under no intervention for covariate means and risks.[41] To assess the sensitivity of the estimates regarding the exclusion of participants with missing data due to the earliest residential address record after the year 2001, we assumed the residential address remained unchanged from the beginning of the year 2001 to the earliest record date for these participants. To be comparable with previous studies, we also applied the conventional Cox model method with age as the time scale to assess the association between the time-varying 5-year moving average of $PM_{2.5}$ and all-cause dementia, adjusting for sex, educational attainment, and time-varying variables of 5-year moving average of mean temperature, annual population density, and annual GDHI per head index. The Cox model was implemented using the "survival" R package.[42] All two-sided P-values < 0.05 were considered statistically significant.

Results

Of 502,356 participants in the UK Biobank, we followed 291,495 eligible participants who met the criteria of being age ≥ 55 years, without the diagnosis of dementia, and without missing data for confounders on Jan 1st, 2010 (**Figure S1, Table S1**). During the 10-year follow-up, 4,700 participants had the event of dementia, and 19,417 died without any dementia diagnosis from primary care or hospitalization records. **Table 1** shows the characteristics of eligible participants in the baseline year 2010. The average age was 63.2 (25th – 75th percentiles: 59.8, 66.7) years, 54% of participants were female, 93% of participants were of white race/ethnicity, and the pre-baseline 9-year moving average of $PM_{2.5}$ concentration from 2001 to 2009 was 12.12 (25th – 75th percentiles: 10.64, 13.41) $\mu g/m^3$.

Table 2 shows the estimated 10-year risk of dementia under different hypothetical ambient $PM_{2.5}$ interventions. Compared with no intervention, the estimated 10-year risk difference was -

0.54 per 1000 population (95%CI: -1.00, -0.10) with the threshold of 12 μg/m^3, -1.36 per 1000 population (95%CI: -2.44, -0.25) with the threshold of 10 μg/m^3, and -1.92 per 1000 population (95%CI: -3.39, -0.33) with the threshold of 9 μg/m^3. The average proportion of participants who would have experienced a change in the ambient $PM_{2.5}$ exposure at each interval to reach the threshold of 12 μg/m^3, 10 μg/m^3, and 9 μg/m^3 is 17.50%, 39.86%, and 53.22%, respectively.

The magnitude of the estimated 10-year risk difference of dementia, comparing no intervention and one with a threshold of 10 μg/m^3, is larger among those who are older (aged ≥ 65 years at baseline) compared to those with baseline age below 65 years and in males compared to females, although the differences between the subgroups were not statistically significant ($p > 0.05$) (**Table 3**). The estimates did not vary substantially in the sensitivity analyses (**Figure 1**, numeric results in **Table S4**), with the magnitude of the point estimates in the sensitivity analyses being similar to or slightly larger than those in the main analysis. The hazard ratio obtained from the Cox model for the association between per 1 μg/m^3 time-varying 5-year moving average $PM_{2.5}$ concentrations and dementia was 1.05 (95% CI: 1.03, 1.07). The estimated 10-year risk difference of mortality was -2.66 per 1000 population (95%CI: -4.52, -0.42) with the threshold of 12 μg/m^3, -5.31 per 1000 population (95%CI: -9.56, -0.36) with the threshold of 10 μg/m^3, and -6.57 per 1000 population (95%CI: -12.39, 0.15) with the threshold of 9 μg/m^3, when comparing with no intervention on $PM_{2.5}$ (**Table S5**).

Discussion

We found that hypothetical threshold interventions on ambient $PM_{2.5}$ levels in late life reduced the 10-year risk of dementia among the UK Biobank participants from 2010 to 2019. Compared with no intervention, the decrease in the estimated 10-year risk difference was -1.36 per 1000 population (95%CI: -2.44, -0.25) under the hypothetical ambient $PM_{2.5}$ interventions with the

threshold of 10 μg/m^3 (England's 2040 and Scotland's 2020 targets). The magnitude of the estimated 10-year risk difference was larger among those aged ≥ 65 years than those aged 55 to 64 years at baseline, and among males compared to females, although the differences between the subgroups were not statistically significant. Although we found that hypothetical interventions of ambient $PM_{2.5}$ could also reduce the estimated risk of all-cause mortality, which was a competing event for dementia, the total effect, encompassing both the direct effect of an intervention on the risk of dementia and the indirect effects mediated by increased survival, remained a decreased risk of dementia.

Long-term exposure to $PM_{2.5}$ has been significantly associated with dementia in large cohorts. A 3.2 μg/m^3 increase in the 5-year moving average of $PM_{2.5}$ was found to be associated with a hazard ratio (HR) of 1.060 [95% CI: 1.054, 1.066] for incident dementia in a national cohort of 12 million population aged ≥ 65 years in the United States,[32] and in a cohort of 2 million population aged ≥ 55 years in Ontario, Canada a HR of 1.04 [95% CI: 1.03, 1.05] per 4.8 μg/m^3 increase.[43] The UK Biobank cohort has been used recently to examine the association between long-term $PM_{2.5}$ exposure and dementia. However, due to the limited availability of $PM_{2.5}$ data, only time-fixed $PM_{2.5}$ exposure at baseline was used in the analyses, which assumes that $PM_{2.5}$ levels in the UK remained constant over the study periods.[20, 21] We used the Cox model in the sensitivity analysis to assess the association between time-varying 5-year moving average $PM_{2.5}$ concentrations and dementia (HR = 1.05 [95% CI: 1.03, 1.07] per 1 μg/m^3 increase), which is consistent with adverse association found in the aforementioned large cohort studies.

The Cox proportional hazard model, used in the previous studies, is a conventional method used to assess the associations between long-term $PM_{2.5}$ exposure and dementia, with hazard ratio as the measure of association. From the perspective of public policy decisions, the risk difference calculated by the parametric g-formula is more intuitive than the Cox hazard ratio. First, the parametric g-formula is grounded in the potential outcome framework,[44] which allows

for the direct evaluation of the effects of hypothetical policies.[12] In comparison, the cause-specific hazard ratio derived from the Cox model represents the instantaneous probability of dementia conditional on those who are alive and free of dementia. This built-in selection bias of hazard can lead to bias in terms of causal inference.[38] Second, the risk difference provides an absolute measure of the effect of regulatory intervention, while the Cox hazard ratio is a relative measure. When the Cox hazard ratio is used as a substitute for relative risk in the calculation of attributable fraction, it requires the rare disease assumption. However, even when this assumption holds, the approximation may falter because the magnitude of the Cox hazard ratio is affected by the censoring proportion, whereas the risk difference remains essentially unchanged.[45, 46]

Strengths and Limitations

This study has substantial strengths. First, by using the parametric g-formula, we provide a more direct measurement of the effect of hypothetical $PM_{2.5}$ regulations on the risk of dementia, which may be more useful for policymaking. Second, we used a counterfactual estimand (the total effect) within a causal framework to estimate the effect of an intervention on the risk of dementia through all pathways (including the pathway mediated by death), which is not often considered in dementia research.[23, 47] This study also has limitations. First, we should cautiously interpret the estimates, since the validity of the findings as causal effects using the parametric g-formula method relies on several assumptions, including correct model specification, no unmeasured confounding, no measurement error, and consistency. Similar to conventional regression models, the model misspecification can compromise the validity of the estimates. We evaluated the model specification in the parametric g-formula by comparing the observed risk and covariate means with those estimated by the parametric g-formula under no intervention (natural course). The observed and natural-course estimates showed no large discrepancy, which is a necessary but not sufficient condition for no model misspecification (**Figure S3**).[41]

The assumption of no unmeasured confounders is untestable.[44] The potential confounding introduced by individual physiologic and behavioral factors (e.g., body mass index and smoking habits) is of less concern since we used ambient $PM_{2.5}$ concentrations instead of personal exposure as the exposure of interest, and we adjusted for time-varying neighborhood-level sociodemographic covariates.[31] The consistency assumption (i.e., the versions of treatment are well-defined) may not hold, given that $PM_{2.5}$ from some sources (e.g., traffic, fossil fuels, agriculture, and wildfires) are found to primarily contribute to the association between $PM_{2.5}$ and dementia.[48, 49] We used $PM_{2.5}$ mass as the exposure of interest, but the chemical components of $PM_{2.5}$ in this study might vary over time and space. Thus, the exposure definition may be unclear as both the concentration and components of $PM_{2.5}$ could affect the risk of dementia. As there are currently no consistent findings regarding which sources of $PM_{2.5}$ contribute to dementia,[48, 49] further studies are warranted to explore the effects of different $PM_{2.5}$ sources on dementia, which can support the implementation of source-targeted measures to reduce pollution. Second, the residential address records collected in the UK Biobank are subjected to some degree of inaccuracy. The modeled ambient $PM_{2.5}$ has high performance (R^2 = 0.82) at a fine resolution of 1×1 km^2, and we used residential address change history to assess the time-varying exposure, which mitigates measurement error of exposure. Third, the passive case ascertainment of dementia using electronic health records could increase the possibility of outcome misclassification, which may result in more conservative estimates compared to active case ascertainment.[50] Fourth, the estimates of this study may not be generalizable to the general UK population. Only around 5.5% of the invitees participated in the baseline survey of the UK Biobank. The estimates are standardized by the conditional distribution of confounders, while the distribution of confounders in the general population may differ from that of our sample population. For example, the responders were more likely to be female and to live in less socioeconomically deprived areas than the nonresponders.[51]

Conclusions

The application of parametric g-formula offers a direct measurement of the health effects of regulatory policies. Although the estimates might not have a causal interpretation given the potential non-fulfillment of the underlying assumptions, our results suggest that policies reducing ambient $PM_{2.5}$ concentrations may benefit health by lowering the risk of dementia among UK Biobank participants who would experience the $PM_{2.5}$ intervention in late life.

Acknowledgments:

This research has been conducted using the UK Biobank Resource under application number 87264. Current Research Ethics Committee (REC) approval number (2021, 5-year cycle): 21/NW/0157. The access to UK Biobank dataset can be applied at https://ukbiobank.ac.uk/register-apply/. This study uses data provided by patients and collected by the NHS as part of their care and support.

References

1. Weuve J, Bennett EE, Ranker L, et al. Exposure to air pollution in relation to risk of dementia and related outcomes: An updated systematic review of the epidemiological literature. *Environmental Health Perspectives*. Sep 2021;129(9):96001. doi:10.1289/ehp8716

2. Morgan Todd E, Davis David A, Iwata N, et al. Glutamatergic neurons in rodent models respond to nanoscale particulate urban air pollutants in vivo and in vitro. *Environmental Health Perspectives*. 2011/07/01 2011;119(7):1003-1009. doi:10.1289/ehp.1002973

3. Levesque S, Surace MJ, McDonald J, Block ML. Air pollution & the brain: Subchronic diesel exhaust exposure causes neuroinflammation and elevates early markers of neurodegenerative disease. *Journal of Neuroinflammation*. 2011/08/24 2011;8(1):105. doi:10.1186/1742-2094-8-105

4. Sahu B, Mackos AR, Floden AM, Wold LE, Combs CK. Particulate matter exposure exacerbates Amyloid-β plaque deposition and gliosis in APP/PS1 Mice. *Journal of Alzheimer's Disease*. 2021;80:761-774. doi:10.3233/JAD-200919

5. Bhatt DP, Puig KL, Gorr MW, Wold LE, Combs CK. A pilot study to assess effects of long-term inhalation of airborne particulate matter on early Alzheimer-like changes in the mouse brain. *PLOS ONE*. 2015;10(5):e0127102. doi:10.1371/journal.pone.0127102

6. Cacciottolo M, Wang X, Driscoll I, et al. Particulate air pollutants, APOE alleles and their contributions to cognitive impairment in older women and to amyloidogenesis in experimental models. *Translational Psychiatry*. 2017/01/01 2017;7(1):e1022-e1022. doi:10.1038/tp.2016.280

7. Patten Kelley T, Valenzuela Anthony E, Wallis C, et al. The effects of chronic exposure to ambient traffic-related air pollution on Alzheimer's disease phenotypes in wildtype and genetically predisposed male and female rats. *Environmental Health Perspectives*. 2021;129(5):057005. doi:10.1289/EHP8905

8. Lee S-H, Chen Y-H, Chien C-C, et al. Three month inhalation exposure to low-level PM2.5 induced brain toxicity in an Alzheimer's disease mouse model. *PLOS ONE*. 2021;16(8):e0254587. doi:10.1371/journal.pone.0254587

9. Turnock ST, Butt EW, Richardson TB, et al. The impact of European legislative and technology measures to reduce air pollutants on air quality, human health and climate. *Environmental Research Letters*. 2016/02/12 2016;11(2):024010. doi:10.1088/1748-9326/11/2/024010

10. Crippa M, Janssens-Maenhout G, Dentener F, et al. Forty years of improvements in European air quality: regional policy-industry interactions with global impacts. *Atmos Chem Phys*. 2016;16(6):3825-3841. doi:10.5194/acp-16-3825-2016

11. Hubbell BJ, Crume RV, Evarts DM, Cohen JM. Policy Monitor: Regulation and Progress under the 1990 Clean Air Act Amendments. *Review of Environmental Economics and Policy*. 2010/01/01 2010;4(1):122-138. doi:10.1093/reep/rep019

12. Zigler CM, Dominici F. Point: Clarifying Policy Evidence With Potential-Outcomes Thinking—Beyond Exposure-Response Estimation in Air Pollution Epidemiology. *American Journal of Epidemiology*. 2014;180(12):1133-1140. doi:10.1093/aje/kwu263

13. Cole SR, Richardson DB, Chu H, Naimi AI. Analysis of occupational asbestos exposure and lung cancer mortality using the g formula. *American Journal of Epidemiology*. 2013;177(9):989-996. doi:10.1093/aje/kws343

14. Edwards JK, McGrath LJ, Buckley JP, Schubauer-Berigan MK, Cole SR, Richardson DB. Occupational radon exposure and lung cancer mortality: estimating intervention effects using the parametric g-formula. *Epidemiology*. 2014;25(6)

15. Keil Alexander P, Richardson David B. Reassessing the link between airborne arsenic exposure among anaconda copper smelter workers and multiple causes of death using the parametric g-formula. *Environmental Health Perspectives*. 2017/04/01 2017;125(4):608-614. doi:10.1289/EHP438

16. Lajous M, Willett WC, Robins J, et al. Changes in fish consumption in midlife and the risk of coronary heart disease in men and women. *American Journal of Epidemiology*. 2013;178(3):382-391. doi:10.1093/aje/kws478

17. Danaei G, Pan A, Hu FB, Hernán MA. Hypothetical midlife interventions in women and risk of type 2 diabetes. *Epidemiology*. 2013;24(1)

18. Chen H, Quick M, Kaufman JS, et al. Impact of lowering fine particulate matter from major emission sources on mortality in Canada: A nationwide causal analysis. *Proceedings of the National Academy of Sciences*. 2022/12/06 2022;119(49):e2209490119. doi:10.1073/pnas.2209490119

19. Chen C, Chen H, van Donkelaar A, et al. Using parametric g-computation to estimate the effect of long-term exposure to air pollution on mortality risk and simulate the benefits of hypothetical policies: The Canadian community health survey cohort (2005 to 2015). *Environmental Health Perspectives*. 2023;131(3):037010. doi:10.1289/EHP11095

20. Parra KL, Alexander GE, Raichlen DA, Klimentidis YC, Furlong MA. Exposure to air pollution and risk of incident dementia in the UK Biobank. *Environmental Research*. 2022/06/01/ 2022;209:112895. doi:https://doi.org/10.1016/j.envres.2022.112895

21. Ma H, Li X, Zhou T, Wang M, Heianza Y, Qi L. Long-term exposure to low-level air pollution, genetic susceptibility and risk of dementia. *International Journal of Epidemiology*. 2023;52(3):738-748. doi:10.1093/ije/dyac146

22. Wang J, Gao Y, Lin Y, et al. Interactive effects of air pollutants and temperature on incidence of dementia: a prospective cohort study. *Environmental Research Letters*. 2023/07/04 2023;18(7):074034. doi:10.1088/1748-9326/acdf73

23. Rojas-Saunero LP, Young JG, Didelez V, Ikram MA, Swanson SA. Considering questions before methods in dementia research with competing events and causal goals. *American Journal of Epidemiology*. 2023;192(8):1415-1423. doi:10.1093/aje/kwad090

24. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*. 2015;12(3):e1001779.

25. UK Biobank. Address change history derivation. 2024. Accessed 01/01/2024, https://biobank.ndph.ox.ac.uk/crystal/refer.cgi?id=526

26. Liu R, Ma Z, Gasparrini A, de la Cruz A, Bi J, Chen K. Integrating augmented in situ measurements and a spatiotemporal machine learning model to back extrapolate historical particulate matter pollution over the United Kingdom: 1980–2019. *Environmental Science & Technology*. 2023/12/12 2023;doi:10.1021/acs.est.3c05424

27. UK Biobank. Environmental Exposures Metadata. Accessed 01/01/2024, https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=2010

28. Hersbach H, Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. ERA5 monthly averaged data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Accessed 07/07/2024, https://doi.org/10.24381/cds.f17050d7

29. Oguz S, Knight J. Regional economic indicators: with a focus on gross disposable household income. *Economic & Labour Market Review*. 2010/05/01 2010;4(5):51-66. doi:10.1057/elmr.2010.68

30. Zhang Y, Chen S-D, Deng Y-T, et al. Identifying modifiable factors and their joint effect on dementia risk in the UK Biobank. *Nature Human Behaviour*. 2023/07/01 2023;7(7):1185-1195. doi:10.1038/s41562-023-01585-x

31. Weisskopf MG, Webster TF. Trade-offs of personal versus more proxy exposure measures in environmental epidemiology. *Epidemiology*. 2017;28(5)

32. Shi L, Steenland K, Li H, et al. A national cohort study (2000–2018) of long-term air pollution exposure and incident dementia in older adults in the United States. *Nature Communications*. 2021/11/19 2021;12(1):6754. doi:10.1038/s41467-021-27049-2

33. Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001;88(4):907-919. doi:10.1093/biomet/88.4.907

34. Wilkinson T, Schnier C, Bush K, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *European Journal of Epidemiology*. 2019/06/01 2019;34(6):557-565. doi:10.1007/s10654-019-00499-1

35. Young JG, Hernán MA, Robins JM. Identification, Estimation and Approximation of Risk under Interventions that Depend on the Natural Value of Treatment Using Observational Data. *Epidemiologic Methods*. 2014;3(1):1-19. doi:doi:10.1515/em-2012-0001

36. Air quality statistics in the UK, 1987 to 2022 - Particulate matter (PM10/PM2.5). Accessed 01/01/2024, https://www.gov.uk/government/statistics/air-quality-statistics/concentrations-of-particulate-matter-pm10-and-pm25

37. National Ambient Air Quality Standards (NAAQS) for PM. Accessed 01/01/2024, https://www.epa.gov/pm-pollution/national-ambient-air-quality-standards-naaqs-pm

38. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in Medicine*. 2020/04/15 2020;39(8):1199-1236. doi:https://doi.org/10.1002/sim.8471

39. McGrath S, Lin V, Zhang Z, et al. gfoRmula: an R package for estimating the effects of sustained treatment strategies via the parametric g-formula. *Patterns*. 2020/06/12/ 2020;1(3):100008. doi:https://doi.org/10.1016/j.patter.2020.100008

40. Mukadam N, Marston L, Lewis G, Mathur R, Rait G, Livingston G. Incidence, age at diagnosis and survival with dementia across ethnic groups in England: A longitudinal study using electronic health records. *Alzheimer's & Dementia*. 2023/04/01 2023;19(4):1300-1307. doi:https://doi.org/10.1002/alz.12774

41. Chiu Y-H, Wen L, McGrath S, Logan R, Dahabreh IJ, Hernán MA. Evaluating model specification when using the parametric g-formula in the presence of censoring. *American Journal of Epidemiology*. 2023:kwad143. doi:10.1093/aje/kwad143

42. Therneau TM, Grambsch PM, Therneau TM, Grambsch PM. *The Cox model*. Springer; 2000.

43. Chen H, Kwong JC, Copes R, et al. Exposure to ambient air pollution and the incidence of dementia: A population-based cohort study. *Environment International*. 2017/11/01/ 2017;108:271-277. doi:https://doi.org/10.1016/j.envint.2017.08.020

44. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986/01/01/ 1986;7(9):1393-1512. doi:https://doi.org/10.1016/0270-0255(86)90088-6

45. Stensrud MJ, Hernán MA. Why Test for Proportional Hazards? *JAMA*. 2020;323(14):1401-1402. doi:10.1001/jama.2020.1267

46. Greenland S. Concepts and pitfalls in measuring and interpreting attributable fractions, prevented fractions, and causation probabilities. *Annals of Epidemiology*. 2015/03/01/ 2015;25(3):155-161. doi:https://doi.org/10.1016/j.annepidem.2014.11.005

47. Letellier N, Gutierrez L-A, Duchesne J, et al. Air quality improvement and incident dementia: Effects of observed and hypothetical reductions in air pollutant using parametric g-computation. *Alzheimer's & Dementia*. 2022/12/01 2022;18(12):2509-2517. doi:https://doi.org/10.1002/alz.12606

48. Shi L, Zhu Q, Wang Y, et al. Incident dementia and long-term exposure to constituents of fine particle air pollution: A national cohort study in the United States. *Proceedings of the National Academy of Sciences*. 2023/01/03 2023;120(1):e2211282119. doi:10.1073/pnas.2211282119

49. Zhang B, Weuve J, Langa KM, et al. Comparison of Particulate Air Pollution From Different Emission Sources and Incident Dementia in the US. *JAMA Internal Medicine*. 2023;183(10):1080-1089. doi:10.1001/jamainternmed.2023.3300

50. Wilker EH, Osman M, Weisskopf MG. Ambient air pollution and clinical dementia: systematic review and meta-analysis. *BMJ*. 2023;381:e071620. doi:10.1136/bmj-2022-071620

51. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

Table 1. Baseline characteristics of eligible participants in the UK Biobank at enrollment, 2006-2010.

Characteristic	n = 291,495
Age in the year 2010, mean (25th – 75th percentiles), y	63.2 (59.8, 66.7)
Sex, No. (%)	
Female	156,929 (54%)
Male	134,566 (46%)
Educational attainment, No. (%)	
College degree	83,764 (29%)
Without college degree	201,596 (69%)
Unknown[a]	6,135 (2.1%)
Race/Ethnicity, No. (%)	
White	269,936 (93%)
Mixed	8,734 (3.0%)
Asian or Asian British	7,923 (2.7%)
Black or Black British	1,016 (0.4%)
Chinese	586 (0.2%)
Other ethnic group	1,547 (0.5%)
Missing	1,753
9-year (2001-2009) moving average of annual mean temperature, mean (25th – 75th percentiles), °C	10.11 (9.58, 10.65)
9-year (2001-2009) moving average of GDHI per head index, mean (25th – 75th percentiles)	98 (82, 107)
9-year (2001-2009) moving average of population density, mean (25th – 75th percentiles), person/m²	0.19 (0.05, 0.22)
9-year (2001-2009) moving average of $PM_{2.5}$ concentration, mean (25th – 75th percentiles), µg/m³	12.12 (10.64, 13.41)
9-year (2001-2009) moving average of warm season (June to August) mean temperature, mean (25th – 75th percentiles), °C	15.75 (15.15, 16.30)
9-year (2001-2009) moving average of cold season (December to February) mean temperature, mean (25th – 75th percentiles), °C	4.77 (4.44, 5.12)

Characteristic	n = 291,495
Modeled annual average exposure to nitrogen dioxide for the year 2010, mean (25th – 75th percentiles), µg/m³ [b]	26 (21, 31)
Missing	3,376
Modeled annual average exposure to nitrogen oxides for the year 2010, mean (25th – 75th percentiles), µg/m³ [b]	43 (34, 50)
Missing	3,376
Average total household income before tax, No. (%) [b]	
Less than 18,000£	66,227 (23%)
18,000£ to 30,999£	71,295 (25%)
31,000£ to 51,999£	57,437 (20%)
52,000£ to 100,000£	36,154 (13%)
Greater than 100,000£	8,714 (3.0%)
Don't know	14,223 (4.9%)
Unknown[a]	37,445 (13%)

Abbreviations: GDHI, gross disposable household income; $PM_{2.5}$, particulate matter with an aerodynamic diameter less than or equal to 2.5µm

[a] The "Unknown" category comprises individuals who either selected "prefer not to answer" or had missing data.

[b] The modeled annual average exposure to nitrogen dioxide and nitrogen oxides for the year 2010, and the average total household income before tax, were additionally adjusted in the sensitivity analyses. In the sensitivity analysis that included additional adjustments for modeled nitrogen dioxide and nitrogen oxides, participants with missing data for these variables were excluded from the analysis.

Table 2. Estimated 10-year risk (95% CI) of dementia under hypothetical ambient $PM_{2.5}$ interventions (achieving annual average standards of 12 μg/m^3, 10 μg/m^3, and 9 μg/m^3) among eligible participants (n = 291,495) in the UK Biobank, 2010-2019. The risk difference is the absolute difference in the 10-year risk when comparing the hypothetical $PM_{2.5}$ intervention with no intervention on $PM_{2.5}$.

Hypothetical intervention	10-year risk (95% CI), ‰	Risk difference (95% CI), ‰	Average % intervened on[a]
No intervention	16.49 (15.79, 16.90)	0 (reference)	—
Threshold			
≤ 12 μg/m³	15.95 (15.08, 16.55)	-0.54 (-1.00, -0.10)	17.50
≤ 10 μg/m³	15.13 (13.86, 16.36)	-1.36 (-2.44, -0.25)	39.86
≤ 9 μg/m³	14.57 (12.93, 16.16)	-1.92 (-3.39, -0.33)	53.22

Note: We used the parametric g-formula to estimate the 10-year risk of dementia adjusting for baseline covariates: sex, baseline age, educational attainment, 9-year moving average (2001-2009) of annual mean temperature (°C), 9-year moving average (2001-2009) of population density (person/m²), 9-year moving average (2001-2009) of GDHI per head index, and 9-year moving average (2001-2009) of $PM_{2.5}$ (μg/m³); and time-varying covariates: annual mean temperature (°C), annual population density (person/m²), annual GDHI per head index, and annual mean $PM_{2.5}$ (μg/m³).

[a] Average proportion of individuals who experienced intervention on ambient $PM_{2.5}$ in each time period to adhere to the hypothetical $PM_{2.5}$ intervention.

Abbreviations: GDHI, gross disposable household income; $PM_{2.5}$, particulate matter with an aerodynamic diameter less than or equal to 2.5μm; ‰, per mille.

Table 3. Estimated 10-year risk of dementia (95% CI) under no intervention, and a hypothetical ambient $PM_{2.5}$ intervention (threshold value of 10 μg/m³), by subgroups among eligible participants in the UK Biobank, 2010-2019 (n = 291,495). The risk difference is the absolute difference in the 10-year risk when comparing the hypothetical $PM_{2.5}$ intervention with no intervention on $PM_{2.5}$.

Subgroup	10-year risk (95% CI), ‰		Risk difference (95% CI), ‰	P-value for subgroup difference
	No intervention	≤ 10 μg/m³		
Female (n = 156,929)	14.91 (13.78, 15.48)	14.42 (12.24, 16.18)	-0.50 (-2.14, 1.23)	0.13
Male (n = 134,566)	18.46 (17.71, 19.20)	16.04 (14.10, 17.94)	-2.41 (-4.05, -0.49)	
Age 55-64 years at baseline (n = 183,659)	7.35 (6.92, 7.71)	6.53 (5.60, 7.37)	-0.82 (-1.69, 0.14)	0.26
Age ≥ 65 years at baseline (n = 107,836)	31.14 (30.10, 32.27)	28.75 (26.08, 31.48)	-2.40 (-5.06, 0.71)	

Note: We used the parametric g-formula to estimate the 10-year risk of dementia adjusting for baseline covariates: sex (for subgroup by age), baseline age, educational attainment, 9-year moving average (2001-2009) of annual mean temperature (°C), 9-year moving average (2001-2009) of population density (person/m²), 9-year moving average (2001-2009) of GDHI per head index, and 9-year moving average (2001-2009) of $PM_{2.5}$ (μg/m³); and time-varying covariates: annual mean temperature (°C), annual population density (person/m²), annual GDHI per head index, and annual mean $PM_{2.5}$ (μg/m³).

Abbreviations: GDHI, gross disposable household income; $PM_{2.5}$, particulate matter with an aerodynamic diameter less than or equal to 2.5μm; ‰, per mille.

Figure 1:

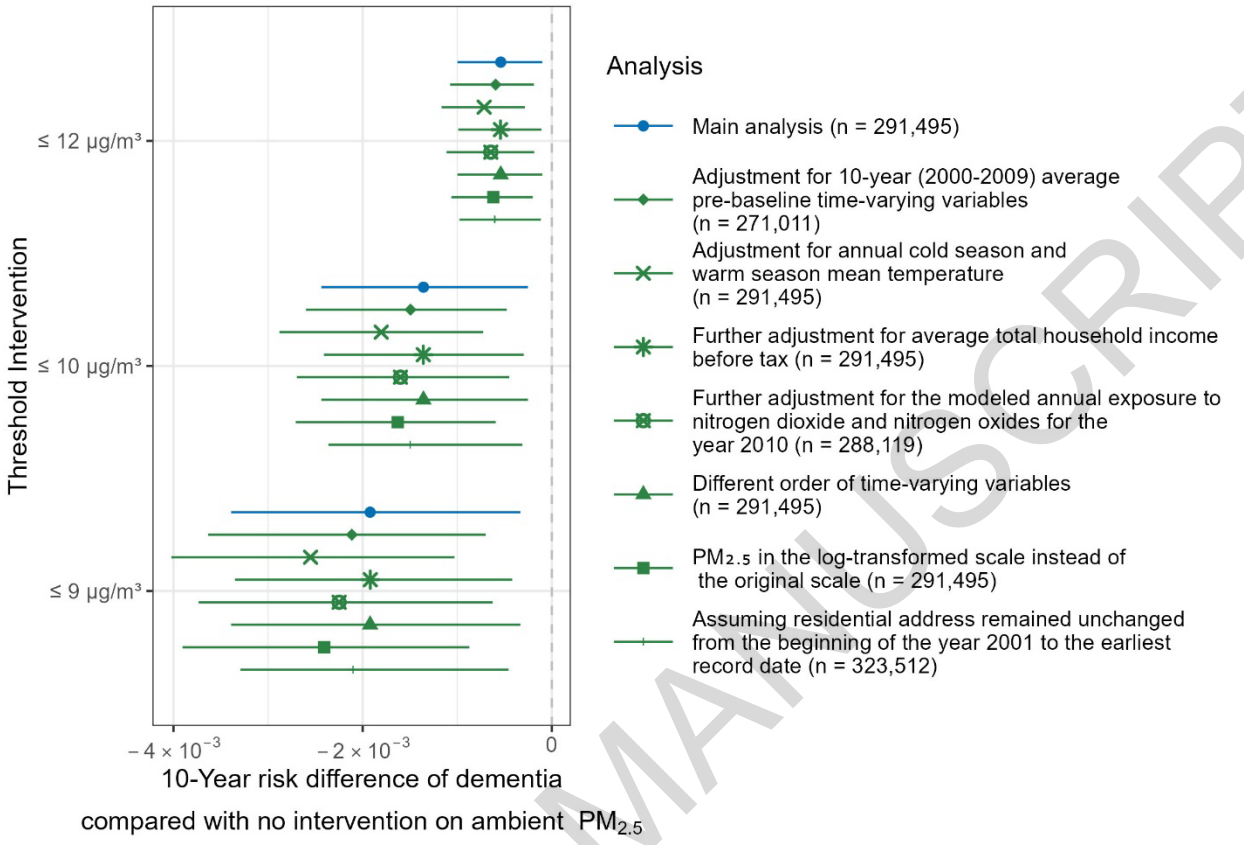

Figure 1. Estimated 10-year risk difference (95% CI) of dementia between different threshold interventions on ambient $PM_{2.5}$ and no intervention for the main analysis and sensitivity analyses. The risk difference is the absolute difference in the 10-year risk when comparing the hypothetical $PM_{2.5}$ intervention with no intervention on $PM_{2.5}$.

Notes: In the main analysis, we used the parametric g-formula to estimate the 10-year risk of dementia adjusting for baseline covariates: sex, baseline age, educational attainment, 9-year moving average (2001-2009) of annual mean temperature (°C), 9-year moving average (2001-2009) of population density (person/m²), 9-year moving average (2001-2009) of GDHI per head index, and 9-year moving average (2001-2009) of $PM_{2.5}$ (µg/m³); and time-varying covariates: annual mean temperature (°C), annual population density (person/m²), annual GDHI per head index, and annual mean $PM_{2.5}$ (µg/m³). The sensitivity analyses are shown in green below the main analyses, with differing factors described in the figure. Numeric results presented in Table S4.

Abbreviations: GDHI, gross disposable household income; $PM_{2.5}$, particulate matter with an aerodynamic diameter less than or equal to 2.5µm.