medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Long-term air pollution exposure and diabetes risk in the elderly population

Running title: air pollution and diabetes

Maayan Yitshak Sade, Ph.D. [1], Liuhua Shi, Sc.D. [2], Elena Colicino, Ph.D. [1], Heresh Amini, Ph.D. [3], Joel D Schwartz, Ph.D. [4], Qian Di, Ph.D. [5], and Robert O Wright, MD MPH. [1].

[1] Icahn School of Medicine at Mount Sinai, Department of Environmental Medicine and Public Health, New York, NY, USA.

[2] Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA.

[3] Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[4] Exposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[5] Vanke School of Public Health, Tsinghua University, Beijing, China.

Corresponding author:

Maayan Yitshak Sade

Department of Environmental Medicine and Public Health

Icahn School of Medicine at Mount Sinai

New York, New York, United States

maayan.yitshak-sade@mssm.edu

Keywords: air pollution; diabetes; particulate matter, ozone, nitrogen dioxide.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

Objective: Type 2 diabetes mellitus is a major public health concern. We assessed the association between air pollution and first documented diabetes occurrence in a national U.S. cohort of Medicare enrollees to estimate diabetes risk.

Research design and methods: We included all Medicare enrollees 65 years and older in the fee-for-service program, part A and part B, in the contiguous United States (2000-2016). Participants were followed annually until the first recorded diabetes diagnosis, end of enrollment, or death. We obtained air pollution annual estimates of fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), and warm-months ozone (O_3) exposures from highly spatiotemporally resolved prediction models. We assessed the simultaneous effect of the pollutants on diabetes risk using Poisson survival analysis with adjustment for temporal and spatial confounders. We repeated the models in data restricted to ZIP codes with air pollution levels not exceeding the ambient air quality standards during the study period.

Results: We have included 264,869,458 person-years of 41,780,637 people. We observed nonlinear associations between the three pollutants and diabetes, with larger risks at lower levels for $PM_{2.5}$ and O_3. When restricting the data to lower air pollution levels, an increased risk for diabetes (Incidence Rate Ratio [95% Confidence Interval] was associated with interquartile range (IQR) increases in $PM_{2.5}$ (1.048 [1.045;1.051]), O_3 (1.016 [1.014;1.18]), and NO_2 (1.040 [1.037; 1.043]).

Conclusion: We found increased diabetes risk associated with air pollution exposures. The observed effects remained in exposure levels below the national ambient air quality standards in the U.S.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a major public health concern that is rising rapidly, with the number of people diagnosed with the disease worldwide more than doubling in the past 20 years (1). T2DM may cause major complications, including blindness, cardiovascular damage, and premature mortality (2). Therefore, it is important to identify determinants of the disease, especially environmental exposures, due to the universal extent of these exposures.

Like many chronic conditions, T1DM and T2DM have a genetic component, but genetics alone explains only a small portion of the variance. Studies show that in T2DM, which represents over 99% of diabetes cases globally (3), both genetics and environmental exposures play a significant role (4). In the last few decades, environmental research focused primarily on behavioral factors, such as inactivity and diet. However, in recent years, cardiometabolic risk was linked to various environmental components, including the social environment, environmental pollutants (3), and, more specifically, air pollution. A recent meta-analysis concluded that each 10 $\mu g/m^3$ increase in long-term fine particulate matter ($PM_{2.5}$) exposure was associated with an 8% increase in T2DM prevalence risk, and each 10 $\mu g/m^3$ increase in long-term nitrogen dioxide (NO_2) exposure was associated with a 7% increase in T2DM prevalence. O_3 was not included in that analysis since the number of studies was limited (5).

These positive associations were observed in other systematic reviews as well (6, 7). However, since it is challenging to investigate the effects of the environment on chronic conditions, current research has many limitations. These limitations include potential bias in the outcome and exposure assessment, inability to separate incidence and prevalence cases, healthy survivor bias, and the inability to account for cardiometabolic risk factors and co-morbidities (6).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Our study addresses many of these limitations by investigating the association between particulate and gaseous air pollution and first documented diabetes occurrence. In a national analysis, using highly spatiotemporally resolved exposure models and comprehensive chronic conditions data on Medicare enrollees across the U.S., we assessed the air pollution effect on first documented diabetes occurrence. We additionally assessed the effect of air pollution exposure below the U.S. Environmental Protection Agency (EPA) national ambient air quality standards (NAAQS) and explored the role of sex and race in susceptibility to the air pollution effect on diabetes risk.

METHODS

Study population

We included all Medicare enrollees who were 65 years and older in the fee-for-service (FFS) program, part A (hospital insurance) and part B (medical insurance), in the contiguous U.S. between the years 2000-2016. We limited the data to person-years included in these programs because the algorithm used to identify chronic conditions utilizes claims covered by these three programs. We entered participants into the cohort on January 1 of the year after they became Medicare participants for each calendar year until the first recorded diabetes diagnosis, end of enrollment in either of the mentioned Medicare programs or death – whichever came first. To avoid gaps in follow-up, once enrollment in the FFS, Medicare part A or B programs was terminated, those participants were no longer included in the cohort even if they renewed their enrollment in later years.

To better approximate incident cases, we excluded individuals diagnosed with diabetes before or in their first year of enrollment. We obtained information on diabetes status and additional

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

chronic conditions from the chronic conditions warehouse database, which includes pre-defined indicators for 27 chronic conditions. Chronic conditions were identified using an algorithm that incorporates claims indicating that an individual received a service for a specific condition. The algorithm combines inpatient, outpatient, skilled nursing facilities, home health claims, or carrier claims (primarily doctor visits) (8).

This study was approved by the Centers for Medicare & Medicaid Services (CMS) under the data use agreement (#RSCH-2020-55733) and the Institutional Review Board of Emory University (#STUDY00000316), and a waiver of informed consent was granted. The Medicare dataset was stored and analyzed in the Rollins High-Performance Computing (HPC) Cluster at Emory University, with Health Insurance Portability and Accountability Act (HIPAA) compliance.

Exposures

We obtained predictions of $PM_{2.5}$ (24-hour average, $\mu g/m^3$), NO_2 (daily 1-hour maximum, ppb) and ozone (daily maximum of 8-hour average, ppb) exposures from validated prediction models calibrated to measurements at approximately 2000 monitoring stations using an ensemble of three machine learners (neural network, random forest, and gradient boosting) that provided daily estimates for a $1km^2$ grid of the contiguous U.S. (9-11). In brief, each machine learning algorithm incorporated more than 100 predictor variables from satellite data, land-use information, weather data, and chemical transport model simulations. The predictions of the three learners were then integrated using a generalized additive model-based geographically weighted-averaging technique. The model was calibrated using daily pollutants concentrations measured at environmental protection agency (EPA) monitoring sites and demonstrated excellent model performance (average cross-validation R^2=0.89, 0.84, and 0.86 for annual predictions of

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

$PM_{2.5}$, NO_2, and O_3, respectively). To align with the Medicare data, we aggregated gridded exposures to ZIP codes by averaging predictions of grid cells within each ZIP code annually. For ozone, we averaged the exposure only during the warmer months of each year (May – October), a commonly specified time window to examine associations with health outcomes (12).

Covariates

The Medicare data is a dynamic cohort that includes individual-level information on the participants' sex, race, age, Medicaid eligibility, and date of death. We additionally obtained the following ZIP code level covariates from the U.S. Census: median household income, population density, median house value, percent of residents with no high school education, and percent of the population self-identified as black. We obtained data from available U.S. Census years, the annual American Community Survey, and extrapolated values for missing years. We aggregated annual summer and winter temperature ZIP code means from gridded Daymet 1- km models (13).

Statistical analysis

We investigated the effect of the three air pollutants simultaneously using the Anderson-Gill formulation in a mixed Poisson survival analysis (14). We included a random intercept for each ZIP code and adjusted the model for individual-level covariates (age, sex, race, and Medicaid insurance) and ZIP code level time-varying covariates (annual summer and winter mean temperatures, population density, median household income, median house value, percent of residents without a high school diploma, and percent black population). Because the exposures and covariates in our analysis vary annually, we fit a proportional hazard model using the equivalent Poisson regression (15). We generated longitudinal data with a binary Poisson

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

outcome for each subject in the risk set, with a zero value at each non-censored year for all person-years except the year of first diabetes occurrence. To non-parametrically estimate a time-varying baseline hazard, we adjusted the model for a penalized spline function of year (14). The model equation is as follows:

$$\mathrm{LogE}(Y_{i,z})=(\beta_0+Uz)+\beta_1 PM_{z,t}+\beta_2 O_{3z,t}+\beta_3 NO_{2z,t}+\beta_4 P_{i,t}+\beta_5 C_{i,t}+\beta_6 W_{z,t}+s(t).$$

Where $Y_{i,z}$ is the outcome (i.e., diabetes) of person i, in ZIP code z; β_0 is a fixed intercept, and u_z is a random intercept per ZIP code z; $PM_{z,t}$ is the annual mean $PM_{2.5}$ associated with ZIP code z in year t; $O_{3z,t}$ is the warm-season average O_3 associated with ZIP code z in year t; $NO_{2z,t}$ is the annual mean NO_2 associated with ZIP code z in year t; $P_{i,t}$ are the person's individual confounders that do not vary over time (race, and sex); $C_{i,t}$ are the person's individual confounders that do vary over time (age, Medicaid insurance); $W_{z,t}$ are spatial and temporal confounders that vary between ZIP codes and over time (winter and summer mean temperature, median household income, median house value, population density, percent population with no high school diploma, and percent blacks); and *s(t)* is a penalized spline for the year of study.

Secondary analyses

First, to explore the possibility of nonlinear associations, we repeated the models while grouping the exposures into four quartiles and treating the lowest quartile as a reference. Second, to assess the effect of air pollution concentrations below the national ambient air quality standards (NAAQS), we have restricted the data only to ZIP codes in which the air pollution level has not exceeded the annual $PM_{2.5}$ and NO_2 standards during the study period (12 μg/m^3 and 53 ppb respectively) (16). To approximate the O_3 NAAQS, we also restricted the data to ZIP codes in which warm-season O_3 levels did not exceed 70 ppb during the study period.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

To identify potential modification by sex and race, we repeated the model in the restricted sample, including interaction terms between each of the three air pollutants and sex or race. If interaction terms were significant, we ran stratified models to obtain the effect estimates of the three pollutants among men, women, whites, blacks, and other races.

Sensitivity analyses

The exclusion of individuals who are not enrolled in the FFS, part A, and Part B programs can potentially induce a selection bias in the study. Biased results can also occur due to competing mortality risk. To avoid selection bias, we used inverse probability weights. Probabilities of enrollment in the cohort and probability of not dying were modeled, accounting for the subjects' age, sex, race, Medicaid eligibility, and the ZIP code population density, percent population under the poverty line, percent population without a high school diploma, percent black and Hispanic population. We calculated the weight by multiplying the inverse probability of enrolling in the three programs with the inverse probability of being alive. We then repeated the models incorporating the weights.

Results for all the analyses are presented as incidence rate ratios (IRR) and 95% confidence intervals (CI) Per interquartile range (IQR) increase in exposures.

RESULTS

We have included 264,869,458 person-years of 41,780,637 people. The mean age was approximately 76 years, 60% were women, and 90% were white. We observed 10,024,879 diabetes cases (3.8% of person-years) (**Table 1**). The survival time from entering our cohort to diabetes diagnosis ranged from 1 to 17 years, with a median of 3 years. The mean and IQR values of the air pollutants were as follows: $PM_{2.5}$ 10.1 μg/m^3 (4.2 μg/m^3), O_3 43.2 ppb (7.0 ppb), and NO_2 18.9 ppb (13.7 ppb). The correlation between the exposures was low to moderate

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(**Supplementary Table 1**), with the highest correlation observed between $PM_{2.5}$ and NO_2 (r=0.44).

We observed a statistically significant increased risk for diabetes, associated with IQR increase in annual mean $PM_{2.5}$ exposure (IRR 1.005, 95% CI 1.004; 1.006) and IQR increase in annual mean NO_2 exposure (IRR 1.185, 95% CI 1.184; 1.187). Warm-season O_3 exposure was negatively associated with diabetes risk (IRR=0.996, 95% CI 0.995: 0.997). When exploring the possibility of nonlinear associations, we found increased diabetes risk associated with all three air pollutants in the 2nd and 3rd quartiles of exposure (compared to the lowest quartile). For $PM_{2.5}$ and warm-months O_3, exposure in the highest quartile was not associated with increased diabetes risk. For NO_2, we observed larger effects in higher exposure quartiles (**Table 2**).

About 40% of the ZIP codes (15601/38921) never experienced $PM_{2.5}$, NO_2, or O_3 pollution levels higher than the national ambient air quality standards during the study period. As seen in **Figure 1**, ZIP codes excluded in this analysis were mostly located in the mid-west, southeast, and mid-Atlantic U.S. regions. In this restricted data, we still observed a statistically significant increase in diabetes risk associated with IQR increases in annual $PM_{2.5}$ (IRR=1.048, 95% CI 1.045;1.051), warm-months O_3 (IRR=1.016, 95% CI 1.015; 1.018), and annual NO_2 (IRR=1.040, 95% CI 1.037; 1.043) (**Figure 2**).

We found a significant interaction between sex and warm-months O_3 (interaction p value<0.001) or annual NO_2 (interaction p value<0.001), with larger effects among women (IRR=1.022, 95% CI 1.020; 1.024 and IRR=1.045, 95% CI 1.041; 1.049, respectively) compared to men (IRR=1.009, 95% CI 1.007; 1.011 and IRR=1.037, 95% CI 1.032; 1.042, respectively). The interaction with annual $PM_{2.5}$ was not statistically significant (interaction p value=0.140). We also found a significant interaction between the three pollutants and race (p<0.001 for all).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Annual $PM_{2.5}$ effects were stronger among whites, the warm-months O_3 effect was stronger among blacks and other races, and annual NO_2 effects were equally stronger among blacks and whites compared to other races (**Figure 1**). We observed similar trends in stratified analyses in the main analytical data (not restricted to lower air pollution levels) (**Supplementary Table 2**). **Supplementary Table 3** shows the baseline characteristics of those included and excluded from the cohort**.** Person-years included in the cohort were of older people. In addition, the proportion of whites and women was larger, and the proportion of death was much smaller. We conducted a sensitivity analysis and repeated the models, including inverse probability weights, to account for the probability of being enrolled in the cohort and alive. Results of the sensitivity analysis were very similar to the results found in the main analysis (**Supplementary Table 4).**

DISCUSSION

In a large, national cohort, we found evidence of increased risk for diabetes associated with air pollution exposures, even when restricting the data to exposure levels below the national ambient air quality standards set by the U.S. EPA. We found nonlinear associations between the air pollutants and diabetes: $PM_{2.5}$ and warm-months O_3 effects were stronger in lower concentrations, and NO_2 effects were stronger in higher pollution concentrations.

T2DM is characterized by high blood glucose levels and increased insulin resistance leading to vascular damage and metabolic dysfunctions (17). A growing body of literature links ambient air pollution exposure to diabetes risk, as concluded in many systematic reviews in recent years (6, 7). Different mechanisms were proposed to explain the links between air pollution exposure and T2DM. Oxidative stress is considered a major pathway to this association (3). PM exposure is associated with oxidative stress, which can lead to lipid peroxidation, reduction of antioxidants, and activation of pro-inflammatory processes (18, 19). These, in turn, have a major role in the

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

development and progression of metabolic syndrome and diabetes in particular (20, 21). Another suggested pathway is PM-induced mitochondrial dysfunction, which decreases brown adipose tissue (22). Since stimulated brown adipose tissue regulates glucose homeostasis and insulin resistance using glucose and free fatty acids during thermogenesis (23), this may systematically interfere with glucose metabolism and insulin sensitivity. Studies also found indications of oxidative stress and mitochondrial disfunction following NO_2 and O_3 exposure (24-27).

Finally, air pollution exposure may increase insulin resistance through alteration of endothelial function (28). The resulting "endothelial dysfunction" promotes vasodilation, fibrinolysis, and anti-platelet aggregation. Nitric oxide (NO) is a major contributor to the processes that lead to endothelial dysfunction. Chronic hyperglycemia from diabetes leads to an impairment of NO production and activity. Endothelial cells undergo apoptosis and are released into the bloodstream. These changes promote arterial denudation, which triggers atherosclerotic processes (19) and impaired insulin signaling (29).

In agreement with these experimental studies, recent epidemiological studies found associations between air pollution exposure and markers of potential underlying pathways of this association: such as lower insulin sensitivity (30), higher fasting glucose (30, 31), Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) (30), and lipids levels (32).

Our study was focused on three principal air pollutants linked to cardiometabolic health: $PM_{2.5}$, O_3, and NO_2. $PM_{2.5}$ are fine inhalable particles, smaller than 2.5 micrometers, comprising a mixture of solids and liquids. Particulate air pollution is defined by its size, and the toxicity of the particles may vary depending on the source. NO_2 is an air pollutant that originates mostly from traffic and high-temperature combustion processes (33). O_3 at the ground level is formed naturally and following a chemical reaction - where air pollutants emitted from sources such as

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

traffic, industry, and wildfires (i.e., nitrogen oxides and volatile organic compounds) react with sunlight and organic gases, principally from vegetation (34).

The Clean Air Act was last amended in 1990 and requires the U.S. EPA to set national ambient air quality standards that mitigate any harmful consequences of air pollution to human health and the environment (16). An independent committee of experts reviews these standards every five years, determining whether current standards adequately achieve this goal based on the accumulating scientific evidence.

We limited our study sample to ZIP codes with lower air pollution levels to simulate a scenario where air pollution levels never exceeded thresholds considered safe. The major findings of our study are the nonlinear associations and the harmful air pollution effects observed even from levels below the national ambient air quality standards set by the EPA. Paul et al. (35) have found similar dose-response functions when studying the effects of air pollution exposure on diabetes mortality. Like our findings, they observed greater diabetes mortality risks in lower $PM_{2.5}$ and O_3 exposures and higher NO_2 exposure. They additionally found greater $PM_{2.5}$ effects on diabetes incidence risk in lower pollution concentrations. In support of our findings, they too concluded that there is no safe threshold for NO_2 exposure, and even lower exposure levels were associated with increased diabetes mortality risk.

The primary ambient air quality standards aim to protect public health, emphasizing vulnerable populations (16). Our study shows that women are more susceptible to the effects of gaseous air pollutants (i.e., O_3 and NO_2). This higher vulnerability to air pollution-induced diabetes among women was documented in several studies (35, 36). It may be attributed to sex-related physiological differences in inflammatory processes, different lifestyles, and activity habits (36). In some cultures, and populations, women also tend to work closer to home and spend more time

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

at the house, which leads to lower exposure measurement error and smaller bias in the effect estimates (37).

Individuals from racial minorities have been found to be more vulnerable to air pollution cardiometabolic effects in several studies (38, 39). For example, a community-based prospective cohort in the U.S. found worse cardiometabolic effects (i.e., glucose levels, arterial endothelial function, and cardiovascular events) associated with exposure to $PM_{2.5}$ among blacks compared to whites (38). This evidence of racial disparities may be attributed to higher exposure levels among minority populations. The net effect of the long history of housing segregation in the U.S. may be to subject certain populations to disproportionate levels of air pollutants creating health disparities in turn. The significantly higher $PM_{2.5}$ and NO_2 exposures among African American and LatinX subjects compared to white subjects have been documented in several studies (40). In addition, these air pollution disparities can also be further magnified by limited access to health care and resources which is more common in predominantly minority neighborhoods (39).

In our study, non-white populations were more vulnerable to ozone but not to $PM_{2.5}$ or NO_2 exposures. These results are not consistent with other studies which find racial minorities to be more vulnerable to air pollution exposure. In our data, racial minorities were more likely to be excluded for not enrolled in the FFS, part A, and Part B programs (**Supplementary Table 3**). This differential selection into the study by race might have biased the interaction analysis results.

Population differences between people included and excluded from the cohort are the major limitation of this study. This is a limitation of all studies that analyze claims data of the Medicare cohort. To ensure this limitation did not introduce a selection bias into our study, we added a sensitivity analysis and repeated our models, including inverse probability weights. The

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

inference did not change, suggesting that our results were not biased due to differential probabilities of enrollment or death. Another limitation of our study is the computational challenge of analyzing big data, which did not allow us to explore nonlinearity with smoothed functions of exposures. We, therefore, explored nonlinearity by aggregating exposures into quartiles. Since we observed positive, stronger effects in lower concentrations for most exposures, we focused our secondary interaction analyses in ZIP codes with air pollution levels lower than the national ambient air quality standards set by the U.S. EPA. Finally, like other air pollution studies, we might have had exposure misclassification error. However, the use of highly spatiotemporally resolved exposure models reduces this error.

In conclusion, assessing the simultaneous effects of particulate and gaseous air pollutants in a national cohort, we found increased diabetes risk associated with air pollution exposure. The observed effects remained when restricting the data to exposure levels below the national ambient air quality standards. For ozone and $PM_{2.5}$, the effects were even stronger in lower levels of exposure. Finally, we found evidence of larger vulnerability among women. Since current studies of the link between air pollution and diabetes are scarce and often limited in quality or sample size, this national study may add robust evidence important for inferring the causal link between air pollution exposure and the development of diabetes.

ACKNOWLEDGEMENTS

MYS designed the study, performed the analysis, and wrote the paper; LS, HA, QD, and JS developed exposure models and reviewed the paper; EC designed and supervised the analysis; RW revied the paper and supervised the findings of this work.

This study was supported by the HERCULES Center (P30 ES019776), the Mount Sinai transdisciplinary center on early environmental exposures (P30 ES023515), the National Institute

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

on Aging (NIA/NIH R01 AG074357), the National Institute of Environmental Health Sciences (R21 ES032606, R01 ES032242, R01 ES032242, and R01 ES013744, P30 ES000002, R01 ES032418), and the United States Environmental Protection Agency (US EPA) (RD-83587201). Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA. Furthermore, the US EPA does not endorse the purchase of any commercial products or services mentioned in the publication. In addition, during the preparation of this manuscript, EC was supported by the National Institute of Environmental Health Science (NIEHS): R01ES032242, and 5U2CES026555-03. Finally, HA is supported by Novo Nordisk Foundation Challenge Programme: Harnessing the Power of Big Data to Address the Societal Challenge of Aging (NNF17OC0027812).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1. Zimmet PZ, Magliano DJ, Herman WH, Shaw JE. Diabetes: a 21st century challenge. The lancet Diabetes & endocrinology. 2014;2(1):56-64.
2. Papatheodorou K, Papanas N, Banach M, Papazoglou D, Edmonds M. Complications of Diabetes 2016. J Diabetes Res. 2016;2016:6989453.
3. Rajagopalan S, Brook RD. Air pollution and type 2 diabetes: mechanistic insights. Diabetes. 2012;61(12):3037-45.
4. Kaprio J, Tuomilehto J, Koskenvuo M, Romanov K, Reunanen A, Eriksson J, et al. Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. Diabetologia. 1992;35(11):1060-7.
5. Yang BY, Fan S, Thiering E, Seissler J, Nowak D, Dong GH, et al. Ambient air pollution and diabetes: A systematic review and meta-analysis. Environ Res. 2020;180:108817.
6. Eze CI, Hemkens GL, Bucher CH, Hoffmann B, Schindler C, Künzli N, et al. Association between Ambient Air Pollution and Diabetes Mellitus in Europe and North America: Systematic Review and Meta-Analysis. Environmental Health Prospectives. 2015.
7. Liu F, Chen G, Huo W, Wang C, Liu S, Li N, et al. Associations between long-term exposure to ambient air pollution and risk of type 2 diabetes mellitus: A systematic review and meta-analysis. Environ Pollut. 2019;252(Pt B):1235-45.
8. Center for Medicare and Medicaid Services. Chronic condition algorithms 2015 [Available from: https://www2.ccwdata.org/web/guest/condition-categories.
9. Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, et al. Assessing NO(2) Concentration and Model Uncertainty with High Spatiotemporal Resolution across the Contiguous United States Using Ensemble Model Averaging. Environ Sci Technol. 2020;54(3):1372-84.
10. Requia WJ, Di Q, Silvern R, Kelly JT, Koutrakis P, Mickley LJ, et al. An Ensemble Learning Approach for Estimating High Spatiotemporal Resolution of Ground-Level Ozone in the Contiguous United States. Environ Sci Technol. 2020;54(18):11037-47.
11. Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, et al. An ensemble-based model of PM(2.5) concentration across the contiguous United States with high spatiotemporal resolution. Environ Int. 2019;130:104909.
12. Wei Y, Yazdi MD, Di Q, Requia WJ, Dominici F, Zanobetti A, et al. Emulating causal dose-response relations between air pollutants and mortality in the Medicare population. Environ Health. 2021;20(1):53.
13. NASA Distributed Active Archive Center (DAAC). Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3 2020 [Available from: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1328.
14. Whitehead J. Fitting Cox's regression model to survival data using GLIM. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1980;29(3):268-75.
15. Shi L, Wu X, Danesh Yazdi M, Braun D, Abu Awad Y, Wei Y, et al. Long-term effects of $PM_{2.5}$ on neurological disorders in the American Medicare population: a longitudinal cohort study. Lancet Planet Health. 2020;4(12):e557-e65.
16. U.S Environmental Protection Agency. National ambient air quality standards (NAAQS) table. 2021 [Available from: https://www.epa.gov/criteria-air-pollutants/naaqs-table#3.
17. Cooper-DeHoff RM, Pepine CJ. Metabolic syndrome and cardiovascular disease: challenges and opportunities. Clin Cardiol. 2007;30(12):593-7.
18. Birben E, Sahiner UM, Sackesen C, Erzurum S, Kalayci O. Oxidative stress and antioxidant defense. World Allergy Organ J. 2012;5(1):9-19.
19. Brook RD, Rajagopalan S, Pope CA, III, Brook JR, Bhatnagar A, Diez-Roux AV, et al. Particulate Matter Air Pollution and Cardiovascular Disease An Update to the Scientific Statement From the American Heart Association. Circulation. 2010;121(21):2331-78.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

20. Hutcheson R, Rocic P. The metabolic syndrome, oxidative stress, environment, and cardiovascular disease: the great exploration. Exp Diabetes Res. 2012;2012:271028.
21. Lim CC, Thurston GD. Air Pollution, Oxidative Stress, and Diabetes: a Life Course Epidemiologic Perspective. Curr Diab Rep. 2019;19(8):58.
22. Xu Z, Xu X, Zhong M, Hotchkiss IP, Lewandowski RP, Wagner JG, et al. Ambient particulate air pollution induces oxidative stress and alterations of mitochondria and gene expression in brown and white adipose tissues. Particle and fibre toxicology. 2011;8(1):1-14.
23. Cannon B, Nedergaard J. Brown adipose tissue: function and physiological significance. Physiol Rev. 2004;84(1):277-359.
24. Sourij H, Meinitzer A, Pilz S, Grammer TB, Winkelmann BR, Boehm BO, et al. Arginine bioavailability ratios are associated with cardiovascular mortality in patients referred to coronary angiography. Atherosclerosis. 2011;218(1):220-5.
25. Schooneman MG, Vaz FM, Houten SM, Soeters MR. Acylcarnitines: reflecting or inflicting insulin resistance? Diabetes. 2013;62(1):1-8.
26. Zhao YY, Wang HL, Cheng XL, Wei F, Bai X, Lin RC, et al. Metabolomics analysis reveals the association between lipid abnormalities and oxidative stress, inflammation, fibrosis, and Nrf2 dysfunction in aristolochic acid-induced nephropathy. Sci Rep. 2015;5:12936.
27. Alderete TL, Chen Z, Toledo-Corral CM, Contreras ZA, Kim JS, Habre R, et al. Ambient and Traffic-Related Air Pollution Exposures as Novel Risk Factors for Metabolic Dysfunction and Type 2 Diabetes. Curr Epidemiol Rep. 2018;5(2):79-91.
28. Baron A, Steinberg H, Chaker H, Leaming R, Johnson A, Brechtel G. Insulin-mediated skeletal muscle vasodilation contributes to both insulin sensitivity and responsiveness in lean humans. The Journal of clinical investigation. 1995;96(2):786-92.
29. Kampfrath T, Maiseyeu A, Ying Z, Shah Z, Deiuliis JA, Xu X, et al. Chronic fine particulate matter exposure induces systemic vascular dysfunction via NADPH oxidase and TLR4 pathways. Circulation research. 2011;108(6):716-26.
30. Wolf K, Popp A, Schneider A, Breitner S, Hampel R, Rathmann W, et al. Association Between Long-term Exposure to Air Pollution and Biomarkers Related to Insulin Resistance, Subclinical Inflammation, and Adipokines. Diabetes. 2016;65(11):3314-26.
31. Yitshak Sade M, Kloog I, Liberty I, Schwartz J, Novack V. The Association Between Air Pollution Exposure and Glucose and Lipids Levels. The Journal of Clinical Endocrinology & Metabolism 2016.

32. Yitshak Sade M, Kloog I, Liberty IF, Katra I, Novack L, Novack V. Air Pollution and Serum Glucose Levels: A Population-Based Study. Medicine. 2015;94(27):e1093.
33. Lim SS, Vos T, Flaxman AD. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010 (vol 380, pg 2224, 2012). Lancet. 2013;381(9874):1276-.
34. Lange SS, Mulholland SE, Honeycutt ME. What Are the Net Benefits of Reducing the Ozone Standard to 65 ppb? An Alternative Analysis. Int J Environ Res Public Health. 2018;15(8).
35. Paul LA, Burnett RT, Kwong JC, Hystad P, van Donkelaar A, Bai L, et al. The impact of air pollution on the incidence of diabetes and survival among prevalent diabetes cases. Environ Int. 2020;134:105333.
36. Andersen ZJ, Raaschou-Nielsen O, Ketzel M, Jensen SS, Hvidberg M, Loft S, et al. Diabetes incidence and long-term exposure to air pollution: a cohort study. Diabetes care. 2012;35(1):92-8.
37. Brook RD, Jerrett M, Brook JR, Bard RL, Finkelstein MM. The relationship between diabetes mellitus and traffic-related air pollution. Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine. 2008;50(1):32-8.
38. Erqou S, Clougherty JE, Olafiranye O, Magnani JW, Aiyer A, Tripathy S, et al. Particulate Matter Air Pollution and Racial Differences in Cardiovascular Disease Risk. Arterioscler Thromb Vasc Biol. 2018;38(4):935-42.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

39. Yitshak-Sade M, Lane KJ, Fabian MP, Kloog I, Hart JE, Davis B, et al. Race or racial segregation? Modification of the PM2. 5 and cardiovascular mortality association. PloS one. 2020;15(7):e0236479.

40. Bell ML, Zanobetti A, Dominici F. Evidence on Vulnerability and Susceptibility to Health Risks Associated With Short-Term Exposure to Particulate Matter: A Systematic Review and Meta-Analysis. American Journal of Epidemiology. 2013;178(6):865-76.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1. Population characteristics (264,869,458 person years)

Population characteristics	**Summary statistics**
Individual characteristics	
Age, Mean (S.D.)	75.97 (7.7)
Female sex, n (%)	158,138,886 (59.7)
Race, n (%)	
White	238,995,206 (90.2)
Black	14,986,036 (5.7)
Other	10,888,216 (4.1)
Medicaid insurance, n (%)	24,986,082 (9.4)
Diabetes, n (%)	10,024,879 (3.8)
Hypertension, n (%)	96,411,807 (36.4)
Ischemic heart disease, n (%)	57,842,331 (21.8)
Death, n (%)	1,125,608 (0.4)
ZIP code characteristics	
Percent poverty, Mean (SD)	0.1 (0.1)
Population density, Mean (SD)	2,650.1 (7,854.1)
Median house value, Mean (SD)	195,301.5 (159,465.5)
Percent black population, Mean (SD)	0.1 (0.2)
Median household income, Mean (SD)	53,618.0 (22,608.6)
Percent black population, Mean (SD)	0.3 (0.1)

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2. The association between air pollution and diabetes: results of Poisson survival models.

Model	Exposure	IRR (95% CI)	P value
Model 1	$PM_{2.5}$	1.005 (1.004; 1.006)	<0.001
	Warm-months O_3	0.996 (0.995; 0.997)	<0.001
	NO_2	1.185 (1.184; 1.187)	<0.001
Model 2	$PM_{2.5}$		
	Q1 (<8.9 μg/m^3)	Reference	
	Q2 (8.9-10.8 μg/m^3)	1.006 (1.004; 1.007)	<0.001
	Q3 (10.9-13.0 μg/m^3)	1.001 (0.999; 1.003)	0.176
	Q4 (≥13.1 μg/m^3)	0.993 (0.991; 0.996)	<0.001
	Warm-months O_3		
	Q1 (<40 ppb)	Reference	
	Q2 (40-43.4 ppb)	1.053 (1.051; 1.055)	<0.001
	Q3 (43.5-47.2 ppb)	1.030 (1.028; 1.032)	<0.001
	Q4 (≥47.3 ppb)	0.987 (0.985; 0.989)	<0.001
	NO_2		
	Q1 (<12.6 ppb)	Reference	
	Q2 (12.6-19.2 ppb)	1.090 (1.088; 1.092)	<0.001
	Q3 (19.3-28.5 ppb)	1.229 (1.227;1.232)	<0.001
	Q4 (≥28.6 ppb)	1.424 (1.421; 1.428)	<0.001

Table 2 shows the results of two Poisson survival models analyzing the effects of annual fine particulate matter ($PM_{2.5}$), warm-months ozone (O_3), and annual nitrogen dioxide (NO_2) on diabetes risk simultaneously. In **model 1**, we treated the exposures as linear. In **model 2**, we grouped the exposures into quartiles and compared each quartile to the lowest quartile. Each model was adjusted for age, race, sex, Medicaid insurance, annual ZIP code means of summer and winter temperature, and annual ZIP code level sociodemographic variables (population density, percent below poverty line, percent black population, median household income, median house value, and percent population with no high school diploma). Each model also includes a random intercept for each ZIP code and a spline function of year.

IRR=Incidence rate ratio; Q=quartile

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted October 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

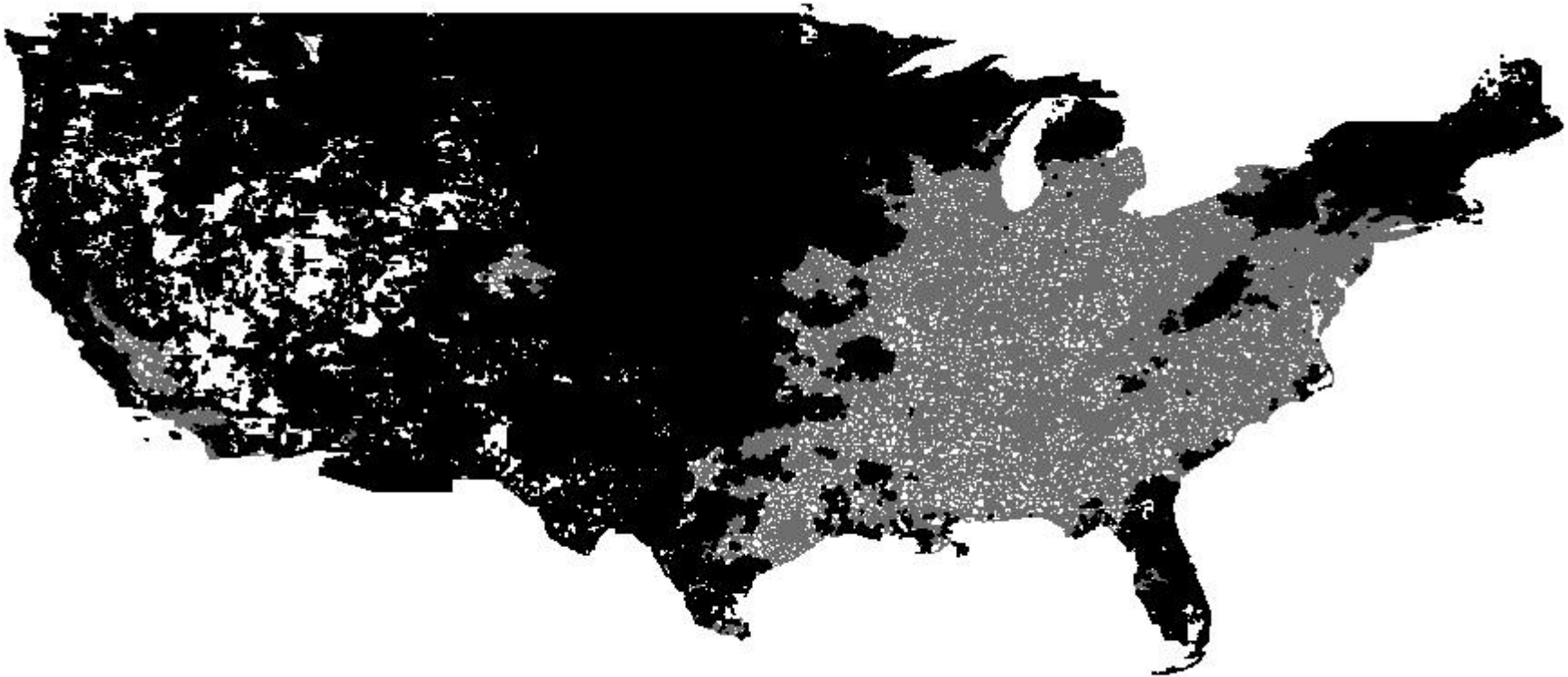
FIGURE LEGENDS

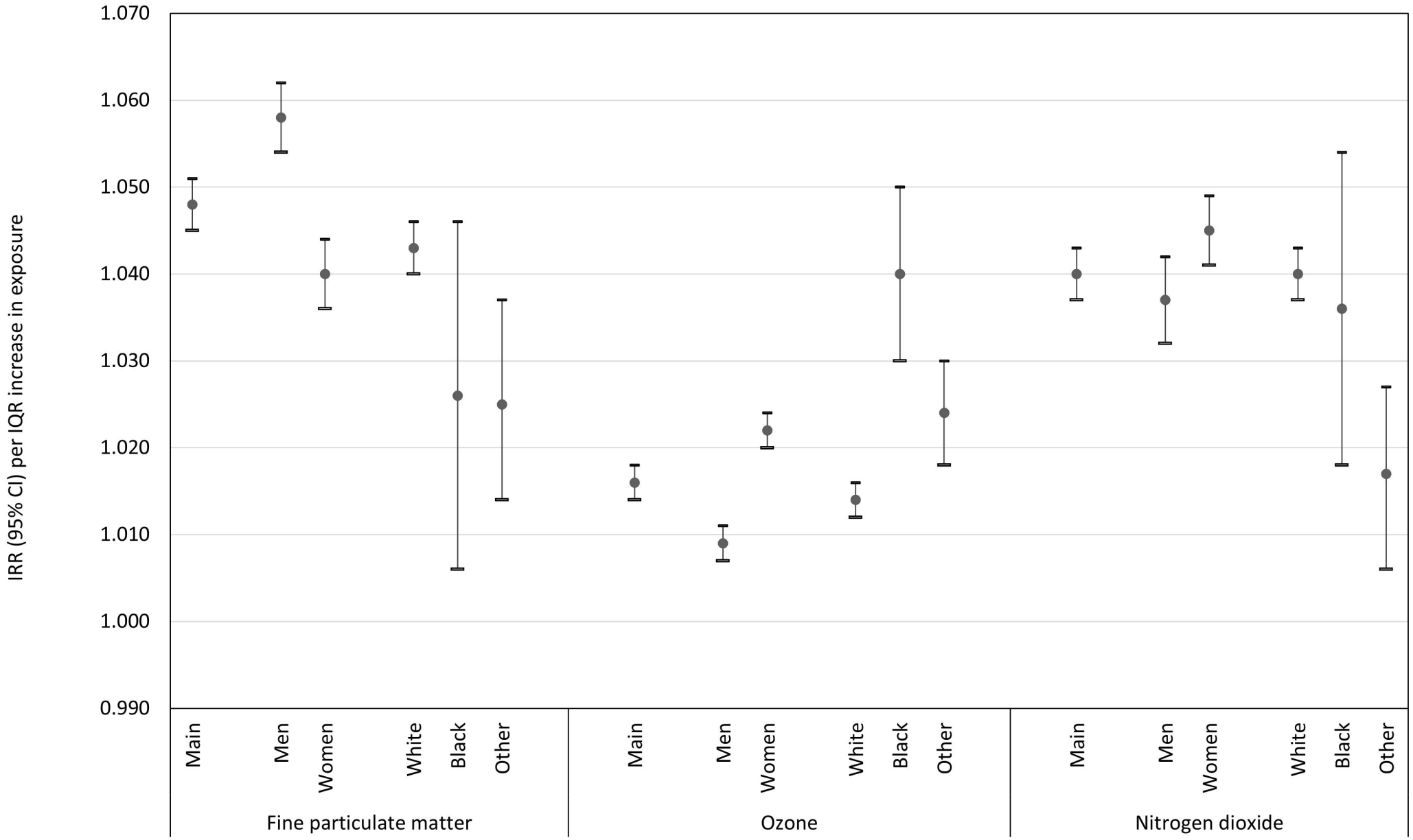
Figure 1. A map of U.S. ZIP codes.

ZIP codes in which annual $PM_{2.5}$, warm-season O_3, and NO_2 levels did not exceed the Environmental Protection Agency's annual national ambient air quality standards (NAAQS) over the study period are colored in black. ZIP codes in which at least one pollutant level exceeded NAAQS in at least one year are colored in gray.

Figure 2. The simultaneous effect of low-levels $PM_{2.5}$, warm-months O_3, and NO_2 exposure on diabetes occurrence.

Figure 2 shows the results of Poisson survival models analyzing the effects of annual fine particulate matter ($PM_{2.5}$), warm-months ozone (O_3), and annual nitrogen dioxide (NO_2) on diabetes occurrence risk simultaneously. All models were restricted to ZIP code in which the annual exposures levels have never exceeded the national ambient air quality standards (i.e., $PM_{2.5}$<12 μg/m^3, NO_2<35 ppb, and O_3<70 ppb). We present the main results and results obtained from models stratified by sex (men or women) and race (white, black, or other). Each model was adjusted for age, Medicaid insurance, annual ZIP code means of summer and winter temperature, and annual ZIP code level sociodemographic variables (population density, percent below poverty line, percent black population, median household income, median house value, and percent population with no high school diploma). Each model also includes a random intercept for each ZIP code and a spline function of year. We added adjustment for sex only in models stratified by race and vice versa.




IRR (95% CI) per IQR increase in exposure
1.070
1.060
1.050
1.040
1.030
1.020
1.010
1.000
0.990
Main
Men
Women
White
Black
Other
Fine particulate matter
Ozone
Nitrogen dioxide