medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 American older adults: a national cohort study.

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

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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 March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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

Background: Type 2 diabetes is a major public health concern. Studies show that both genetics and environmental exposures play a significant role in developing the disease. We assessed the association between air pollution and first documented diabetes occurrence in a national U.S. cohort of older adults to estimate diabetes risk.

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 (264,869,458 person-years). We obtained 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 effects of the pollutants on diabetes risk using survival analyses. We repeated the models in cohorts restricted to ZIP codes with air pollution levels not exceeding the national ambient air quality standards (NAAQS) during the study period.

Results: We identified 10,024,879 diabetes cases of 41,780,637 people (3.8% of person-years). The hazard ratio (HR) for first diabetes occurrence was 1.074 (95% CI 1.058; 1.089) for 5 $\mu g/m^3$ increase in $PM_{2.5}$, 1.055 (95% CI 1.050; 1.060) for 5 ppb increase in NO_2, and 0.999 (95% CI 0.993; 1.004) for 5 ppb increase in O_3. Both for NO_2 and $PM_{2.5}$ there was evidence of non-linear exposure-response curves with stronger associations at lower levels ($NO_2 \leq 40$ ppb, $PM_{2.5} \leq 9$ $\mu g/m^3$). Furthermore, associations remained in the restricted low-level cohorts. The O_3-diabetes exposure-response relationship differed greatly between models and require further investigation.

Conclusion: Exposures to $PM_{2.5}$ and NO_2 are associated with increased diabetes risk, even when restricting the exposure to levels below the NAAQS set by the U.S. EPA.

Keywords: air pollution; diabetes; $PM_{2.5}$, NO_2, O_3

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 (T2D) is a major public health concern rising rapidly, with the number of people diagnosed with the disease worldwide more than doubling in the past 20 years [1]. T2D may cause major complications, including blindness, cardiovascular damage, and premature mortality [2], it is therefore important to identify determinants of the disease.

Like many chronic conditions, type I diabetes (T1D) and T2D have a genetic component, but genetics alone explains only a small portion of the variance. Studies show that in T2D, which represents over 99% of diabetes cases globally [3], 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 [5,6], environmental pollutants [3,7], and, more specifically, air pollution [8-13]. A recent meta-analysis concluded that each ten $\mu g/m^3$ increase in long-term fine particulate matter ($PM_{2.5}$) exposure was associated with an 8% increase in T2D prevalence risk, and each ten $\mu g/m^3$ increase in long-term nitrogen dioxide (NO_2) exposure was associated with a 7% increase in T2D prevalence. O_3 was not included in that analysis since the number of studies was limited [14].

These positive associations were observed in other systematic reviews as well [15,16]. 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 [15].

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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. Our use of highly spatiotemporally resolved exposure models reduces bias due to exposure misclassification. In addition, unlike most national studies or chronic air pollution effects, we use comprehensive chronic conditions data on Medicare enrollees across the U.S. to reduce misclassification of diabetes and approximate incident cases.

In this study, we aim to assess the association between long-term air pollution exposure and first documented diabetes occurrence, focusing on the effect of air pollution exposure below the U.S. Environmental Protection Agency (EPA) national ambient air quality standards (NAAQS).

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 after they became Medicare participants and followed 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.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 better approximate incident cases, we excluded individuals diagnosed with diabetes before or in their first year of enrollment. We obtained information on diabetes status from the chronic conditions warehouse database. Diabetes was identified using an algorithm that incorporates claims indicating that an individual received a healthcare service for diabetes. The algorithm combines inpatient, outpatient, skilled nursing facilities, home health claims, or carrier claims (primarily doctor visits) [17]. To be diagnosed with diabetes, beneficiaries must have at least one claim with the international classification of disease codes of diabetes (ICD-9 or ICD-10) from either inpatient, skilled nursing facility, home health, or part B services within two years.

This study was approved by the Centers for Medicare & Medicaid Services (CMS) under the data use agreement (#RSCH-2020-55733), the Institutional Review Board of Emory University (#STUDY00000316), and the Institutional Review Board of Mount Sinai (STUDY 20-01344), 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, in compliance with Health Insurance Portability and Accountability Act (HIPAA).

Exposures

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. NO_2 is an air pollutant that originates mostly from traffic and high-temperature combustion processes [18]. O_3 at the ground level is formed naturally and following a chemical reaction - where air pollutants emitted from sources such as traffic, industry, and wildfires (i.e., nitrogen oxides and volatile organic compounds) react with sunlight and organic gases, principally from vegetation [19].

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

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 1 km^2 grid of the contiguous U.S. [20-22]. 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 EPA monitoring sites and demonstrated excellent model performance (average cross-validation R^2=0.89, 0.84, and 0.86 for annual predictions of $PM_{2.5}$, 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 [23]. We selected this window because ozone formation is increased in warmer months due to its reaction with sunlight.

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, percent home renters, 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 and the annual American Community Survey, and extrapolated values for missing years. We

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

aggregated annual summer, and winter temperature ZIP code means from gridded Daymet 1- km models [24].

Statistical analysis

We investigated the effect of the three air pollutants simultaneously using a cox-equivalent re-parameterized Poisson survival approach [25]. To allow for strata-specific baseline hazard functions, we stratified the models by follow-up year, calendar year, ZIP code, sex, race (white, black, other), age, and Medicaid insurance. We calculated diabetes counts in each follow-up year, calendar year, and ZIP code within strata specified by these individual characteristics, using the log of the corresponding total person-time as an offset. This approach is proven to be equivalent to a time-varying cox survival model using the Anderson-Gill formulation [25]. We used the m-out-n bootstrap method to account for spatial autocorrelation of neighboring ZIP codes. This methods samples ZIP codes randomly for each bootstrap replicate, and estimates statistically robust confidence intervals [26]. First, we fit a multi-pollutant model using linear terms for the three pollutants. Then, to allow for nonlinear exposure-response curves, we used penalized spline functions for each of the pollutants. We fit three multipollutant models in which we used a penalized spline to estimate the exposure-response curve for one pollutant, and linear terms for the adjusted pollutants. We adjusted the models for annual summer and winter mean temperatures, population density, median household income, percent home renters, percent of residents without a high school diploma, and percent black population.

To assess the effect of air pollution concentrations below the NAAQS, we created three restricted low-exposure subsets comprised of individuals who were always exposed to exposure levels lower than the national standards within the study period (i.e., $PM_{2.5}$ <12 μg/m^3, NO_2 <53 ppb). As there is no annual standard for O_3, we selected 50 ppb as a threshold to approximate the

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

current World Health Organization global interim target 1 for peak-season average O_3 concentration [27].

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 risks. 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 these probabilities using pooled logistic regressions. We then calculated the weight by multiplying the inverse probability of enrolling in the three programs with the inverse probability of being alive and calculated the averaged weight within each stratum of follow up year, calendar year, ZIP code, and individual characteristics. We then repeated the models incorporating the weights and compared the exposure-response curves with and without weights.

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

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

Using linear terms, we observed an increased diabetes risk associated with 5 μg/m^3 increase in $PM_{2.5}$ (HR=1.074, 95% CI 1.058; 1.089) and 5 ppb increase in NO_2 (HR=1.055, 95% CI 1.050; 1.060). No linear association was observed with O_3 (HR 0.999,95% CI 0.993; 1.004). We then used penalized spline functions to allow for nonlinear exposure-response curves and found evidence of nonlinear associations with the three pollutants. Both for NO_2 and $PM_{2.5}$ we observed stronger associations at the lower ends of the pollutants' distributions. For $PM_{2.5}$ the associations were stronger for levels<8.2 μg/m^3 (the 25th percentile), followed by a plateaued association. For NO_2 the associations were stronger for levels < 36 ppb (94th percentile), followed by attenuated associations. The O_3 exposure-response curve was highly nonlinear with mostly negative associations (**Figure 1**). The associations with NO_2 and $PM_{2.5}$ were similar in the single pollutant models and the multivariate model. The O_3 exposure-response curve differed greatly between the single and multipollutant models. Unlike the negative associations seen in the multivariate model, there was no clear trend in the association with the pollutant in the single pollutant model (**Supplementary Figure 1**).

Of the analytical dataset, 33.7% (89,519,417 person-years), 96.4% (255,575,750 person-years), and 54.4% (144,225,079 person-years) were only exposed to low-level annual $PM_{2.5}$, annual NO_2, and warm-season O_3 during the study period, respectively. High-pollution ZIP codes excluded from the restricted datasets were mostly located in the mid-west, southeast, and the mid-Atlantic U.S. regions. We estimated the nonlinear exposure-response curves for the three pollutants in multivariate regressions in these restricted cohorts, and the associations with $PM_{2.5}$ and NO_2 remained similar and significant. For O_3, there was no clear association with diabetes and the direction of the associations differed across the distribution of the pollutant (**Figure 2**).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 2 shows the baseline characteristics of those included and excluded from the analytical dataset**.** Person-years included in the analytical dataset 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 accounting for the probability of being enrolled in the cohort and alive. The results of the two models were very similar (**Supplementary Figure 2).**

DISCUSSION

In a large-scale, national cohort, we observed higher risks for diabetes associated with increases in NO_2 and $PM_{2.5}$ exposures, even when restricting the data to exposure levels below the NAAQS set by the U.S. EPA. Results regarding the association with O_3 were inconclusive.

T2D is characterized by high blood glucose levels and increased insulin resistance leading to vascular damage and metabolic dysfunctions [29]. A growing body of literature links ambient air pollution exposure to diabetes risk, as concluded in many recent systematic reviews [15,16]. Studies also found associations between air pollution exposure and markers of potential underlying pathways of this association: lower insulin sensitivity [30], and higher fasting glucose [30,31]. Different mechanisms were proposed to explain the links between air pollution exposure and T2D. 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 [9,32]. These, in turn, have a major role in the development and progression of metabolic syndrome and diabetes in particular [33,34]. Another suggested pathway is PM-induced mitochondrial dysfunction, which decreases brown adipose tissue [35]. Since stimulated brown adipose tissue regulates glucose homeostasis and insulin resistance using glucose and free fatty acids during thermogenesis [36,37], this may systematically

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

interfere with glucose metabolism and insulin sensitivity. Studies also found indications of oxidative stress and mitochondrial disfunction following NO_2 and O_3 exposure [38-41].

Our study shows evidence of increased diabetes risk associated with long-term $PM_{2.5}$ and NO_2 exposures. Like our findings, a 2019 meta-analysis concluded that the odds for T2D were 1.03 and 1.05 times higher for incremental increases of 10 $\mu g/m^3$ in $PM_{2.5}$ and NO_2 exposures, respectively [16]. Studies assessing the association between O_3 and diabetes are much more limited, and the direction of the observed associations is inconsistent and varies greatly depending on the model covariates [42]. For example, a longitudinal cohort of 13,548 individuals in China found a 5.7% increase in diabetes incidence hazard associated with a ten $\mu g/m^3$ increase in annual O_3 exposure [43]. However, two recent long-term studies did not find a significantly increased diabetes risk associated with O_3 exposure in multipollutant models accounting for particulate matter [44,45] and NO_2 [44] exposures. Moreover, a 2021 review concluded that evidence on the association between O_3 exposure and diabetes is insufficient to infer causality [42]. Our study also showed inconclusive results, possibly related to residual confounding, exposure measurement error, or the complexities of simultaneously estimating the effects of multiple air pollutants.

The Clean Air Act was last amended in 1990 and requires the U.S. EPA to set NAAQS that mitigate any harmful consequences of air pollution to human health and the environment [28]. However, recent studies suggest that these thresholds are insufficient to protect human health. There is evidence of increased mortality risk associated with air pollution levels below the NAAQS [46-51]. In some analyses, the associations observed at the lower exposure distribution range were stronger than the higher range of the exposure distribution [50,51]. This may be attributed to larger measurement error at the rarer higher exposure concentrations that is likely to

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

attenuate the overall effect observed when considering the full ranges of the pollutant exposures in the analysis [52]. Regarding $PM_{2.5}$, it is also possible that the composition of the particles is different for low and high pollution days [46].

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 by the U.S. EPA. The major findings of our study are the nonlinear associations and the harmful air pollution effects observed even from levels below the NAAQS set by the EPA. These results suggest that NO_2, and $PM_{2.5}$ exposures below the NAAQS are associated with increased cardiometabolic risk. Like our findings, Paul et al. [53] also observed associations between moderate exposure concentrations of $PM_{2.5}$ and NO_2 and diabetes mortality and incidence risks. Moreover, they report greater $PM_{2.5}$ effects on diabetes incidence risk in lower pollution concentrations.

Our study has several limitations. First, there are differences between people included and excluded from the cohort. This is a limitation of all studies that analyze claims data of the Medicare cohort. However, our sensitivity analysis suggests that our results were not biased due to differential probabilities of enrollment or death. Second, the Medicare data does not provide information on subtypes of diabetes or individual confounders such as BMI, smoking, physical activity, or lifestyle. However, since exposure is assigned on a ZIP code level, neighborhood factors are more likely to confound the associations in our study. These neighborhood factors were accounted for in our models. Finally, we might have had exposure misclassification errors like other air pollution studies. 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 $PM_{2.5}$ and NO_2 exposure. The

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

observed effects remained when restricting the data to exposure levels below the NAAQS. For ozone, the effects were inconclusive and require further investigation. 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

This study was supported by the HERCULES Center (P30 ES019776), the Mount Sinai transdisciplinary center on early environmental exposures (P30 ES023515), the National Institute on Aging (NIA/NIH R01 AG074357), the National Institute of Environmental Health Sciences (R21 ES032606, R01 ES032242, 5U2CES026555-03, 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. Finally, H.A. 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 March 30, 2022. 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):3037-3045.
4. Kaprio J, Tuomilehto J, Koskenvuo M, 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):1060-1067.
5. Khan SU, Javed Z, Lone AN, et al. Social Vulnerability and Premature Cardiovascular Mortality Among US Counties, 2014 to 2018. *Circulation.* 2021;144(16):1272-1279.
6. Basile Ibrahim B, Barcelona V, Condon EM, Crusto CA, Taylor JY. The Association Between Neighborhood Social Vulnerability and Cardiovascular Health Risk Among Black/African American Women in the InterGEN Study. *Nurs Res.* 2021;70(5S Suppl 1):S3-s12.
7. Dendup T, Feng X, Clingan S, Astell-Burt T. Environmental Risk Factors for Developing Type 2 Diabetes Mellitus: A Systematic Review. *Int J Environ Res Public Health.* 2018;15(1).
8. Brook RD. Cardiovascular effects of air pollution. *Clinical Science.* 2008;115(5-6):175-187.
9. Brook RD, Rajagopalan S, Pope CA, III, et al. Particulate Matter Air Pollution and Cardiovascular Disease An Update to the Scientific Statement From the American Heart Association. *Circulation.* 2010;121(21):2331-2378.
10. Andersen ZJ, Raaschou-Nielsen O, Ketzel M, et al. Diabetes incidence and long-term exposure to air pollution: a cohort study. *Diabetes care.* 2012;35(1):92-98.
11. Chandrabose M, Rachele JN, Gunn L, et al. Built environment and cardio-metabolic health: systematic review and meta-analysis of longitudinal studies. *Obes Rev.* 2019;20(1):41-54.
12. Li Y, Xu L, Shan Z, Teng W, Han C. Association between air pollution and type 2 diabetes: an updated review of the literature. *Ther Adv Endocrinol Metab.* 2019;10:2042018819897046.
13. Park SK, Wang W. Ambient Air Pollution and Type 2 Diabetes Mellitus: A Systematic Review of Epidemiologic Research. *Curr Envir Health Rpt.* 2014;1:275-286.
14. Yang BY, Fan S, Thiering E, et al. Ambient air pollution and diabetes: A systematic review and meta-analysis. *Environ Res.* 2020;180:108817.
15. Eze CI, Hemkens GL, Bucher CH, et al. Association between Ambient Air Pollution and Diabetes Mellitus in Europe and North America: Systematic Review and Meta-Analysis. *Environmental Health Prospectives.* 2015.
16. Liu F, Chen G, Huo W, 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):1235-1245.
17. Center for Medicare & Medicaid services. Chronic condition algorithms. https://www2.ccwdata.org/web/guest/condition-categories. Published 2015. Accessed Sep 15, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

18. 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):1276-1276.
19. 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).
20. Di Q, Amini H, Shi L, 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):1372-1384.
21. Requia WJ, Di Q, Silvern R, 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):11037-11047.
22. Di Q, Amini H, Shi L, 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.
23. Wei Y, Yazdi MD, Di Q, et al. Emulating causal dose-response relations between air pollutants and mortality in the Medicare population. *Environ Health.* 2021;20(1):53.
24. NASA Distributed Active Archive Center (DAAC) at NSIDC (DAAC). Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3. https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1328. Published 2020. Accessed August 19, 2021.
25. Shi L, Wu X, Danesh Yazdi M, 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-e565.
26. Bickel P, Götze F. Selected Works of Willem van Zwet. 2012.
27. Organization WH. *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide.* World Health Organization;2005.
28. United States Environmental Protection Agency. National ambient air quality standards (NAAQS) table. https://www.epa.gov/criteria-air-pollutants/naaqs-table#3. Published 2021. Accessed July 15, 2021.
29. Cooper-DeHoff RM, Pepine CJ. Metabolic syndrome and cardiovascular disease: challenges and opportunities. *Clin Cardiol.* 2007;30(12):593-597.
30. Wolf K, Popp A, Schneider A, et al. Association Between Long-term Exposure to Air Pollution and Biomarkers Related to Insulin Resistance, Subclinical Inflammation, and Adipokines. *Diabetes.* 2016;65(11):3314-3326.
31. Yitshak Sade M, Kloog I, Liberty I, Schwartz J, Novack V. The Association Between Air Pollution Exposure andGlucose and Lipids Levels. *The Journal of Clinical Endocrinology & Metabolism* 2016.
32. Birben E, Sahiner UM, Sackesen C, Erzurum S, Kalayci O. Oxidative stress and antioxidant defense. *World Allergy Organ J.* 2012;5(1):9-19.
33. Hutcheson R, Rocic P. The metabolic syndrome, oxidative stress, environment, and cardiovascular disease: the great exploration. *Exp Diabetes Res.* 2012;2012:271028.
34. Lim CC, Thurston GD. Air Pollution, Oxidative Stress, and Diabetes: a Life Course Epidemiologic Perspective. *Curr Diab Rep.* 2019;19(8):58.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

35. Xu Z, Xu X, Zhong M, 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.
36. Cannon B, Nedergaard J. Brown adipose tissue: function and physiological significance. *Physiol Rev.* 2004;84(1):277-359.
37. Bartelt A, Bruns OT, Reimer R, et al. Brown adipose tissue activity controls triglyceride clearance. *Nat Med.* 2011;17(2):200-205.
38. Sourij H, Meinitzer A, Pilz S, et al. Arginine bioavailability ratios are associated with cardiovascular mortality in patients referred to coronary angiography. *Atherosclerosis.* 2011;218(1):220-225.
39. Schooneman MG, Vaz FM, Houten SM, Soeters MR. Acylcarnitines: reflecting or inflicting insulin resistance? *Diabetes.* 2013;62(1):1-8.
40. Zhao YY, Wang HL, Cheng XL, 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.
41. Alderete TL, Chen Z, Toledo-Corral CM, 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.
42. LaKind JS, Burns CJ, Pottenger LH, Naiman DQ, Goodman JE, Marchitti SA. Does ozone inhalation cause adverse metabolic effects in humans? A systematic review. *Crit Rev Toxicol.* 2021;51(6):467-508.
43. Wang Y, Cao R, Xu Z, et al. Long-term exposure to ozone and diabetes incidence: A longitudinal cohort study in China. *Sci Total Environ.* 2022;816:151634.
44. Jerrett M, Brook R, White LF, et al. Ambient ozone and incident diabetes: a prospective analysis in a large cohort of African American women. *Environment international.* 2017;102:42-47.
45. Li H, Duan D, Xu J, et al. Ambient air pollution and risk of type 2 diabetes in the Chinese. *Environ Sci Pollut Res Int.* 2019;26(16):16261-16273.
46. Shi L, Rosenberg A, Wang Y, et al. Low-Concentration Air Pollution and Mortality in American Older Adults: A National Cohort Analysis (2001-2017). *Environ Sci Technol.* 2021.
47. Danesh Yazdi M, Wang Y, Di Q, et al. Long-Term Association of Air Pollution and Hospital Admissions Among Medicare Participants Using a Doubly Robust Additive Model. *Circulation.* 2021;143(16):1584-1596.
48. Shi L, Zanobetti A, Kloog I, et al. Low-Concentration PM2.5 and Mortality: Estimating Acute and Chronic Effects in a Population-Based Study. *Environ Health Perspect.* 2016;124(1):46-52.
49. Yitshak-Sade M, Lane KJ, Fabian MP, et al. Race or racial segregation? Modification of the PM2. 5 and cardiovascular mortality association. *PloS one.* 2020;15(7):e0236479.
50. Wei Y, Wang Y, Wu X, et al. Causal Effects of Air Pollution on Mortality Rate in Massachusetts. *Am J Epidemiol.* 2020;189(11):1316-1323.
51. Di Q, Dai L, Wang Y, et al. Association of Short-term Exposure to Air Pollution With Mortality in Older Adults. *Jama.* 2017;318(24):2446-2456.
52. Steenland K, Karnes C, Darrow L, Barry V. Attenuation of exposure-response rate ratios at higher exposures: a simulation study focusing on frailty and measurement error. *Epidemiology.* 2015;26(3):395-401.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 .

53. Paul LA, Burnett RT, Kwong JC, et al. The impact of air pollution on the incidence of diabetes and survival among prevalent diabetes cases. *Environ Int.* 2020;134:105333.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.09.21263282; this version posted March 30, 2022. 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 March 30, 2022. 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 .

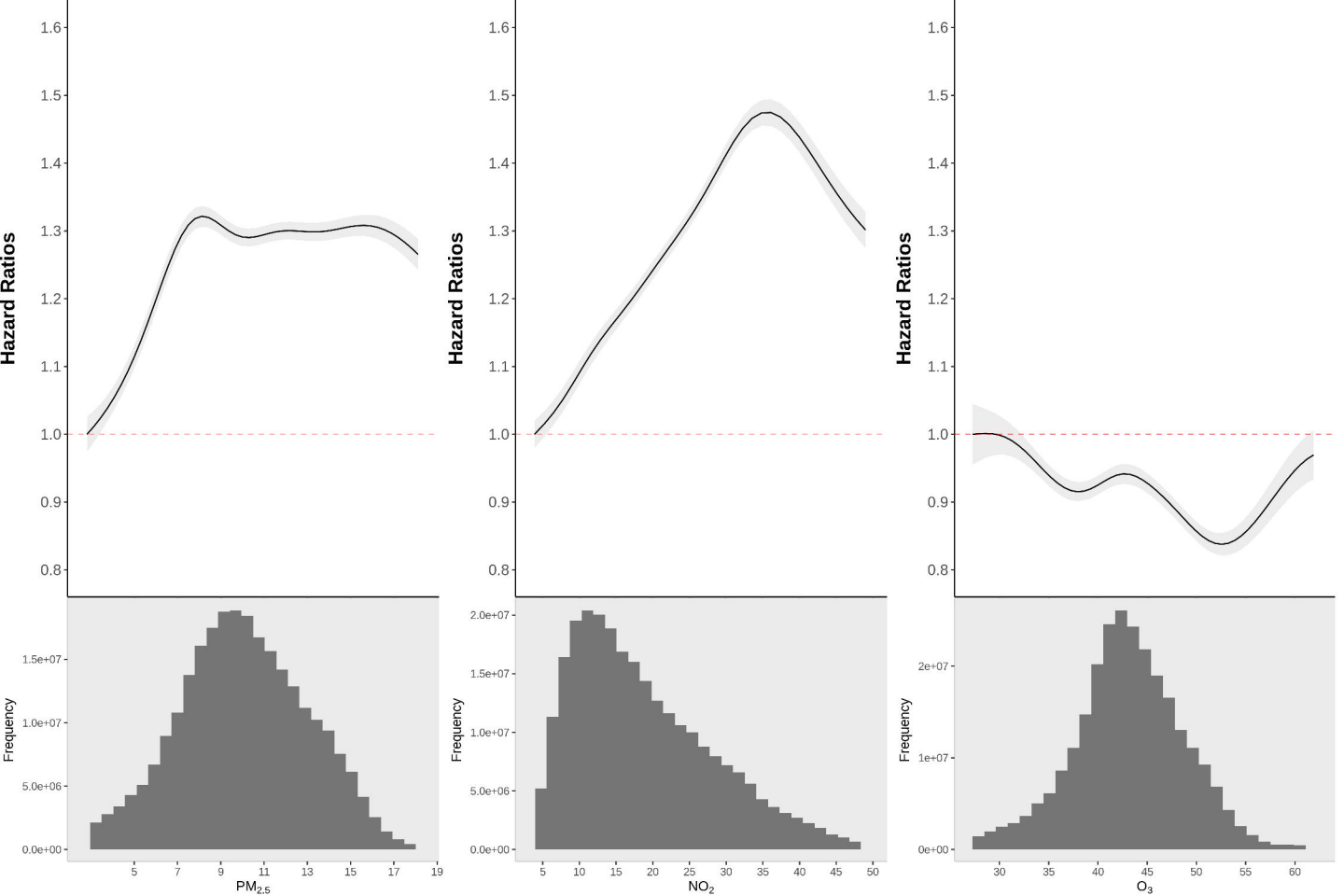
FIGURE LEGENDS

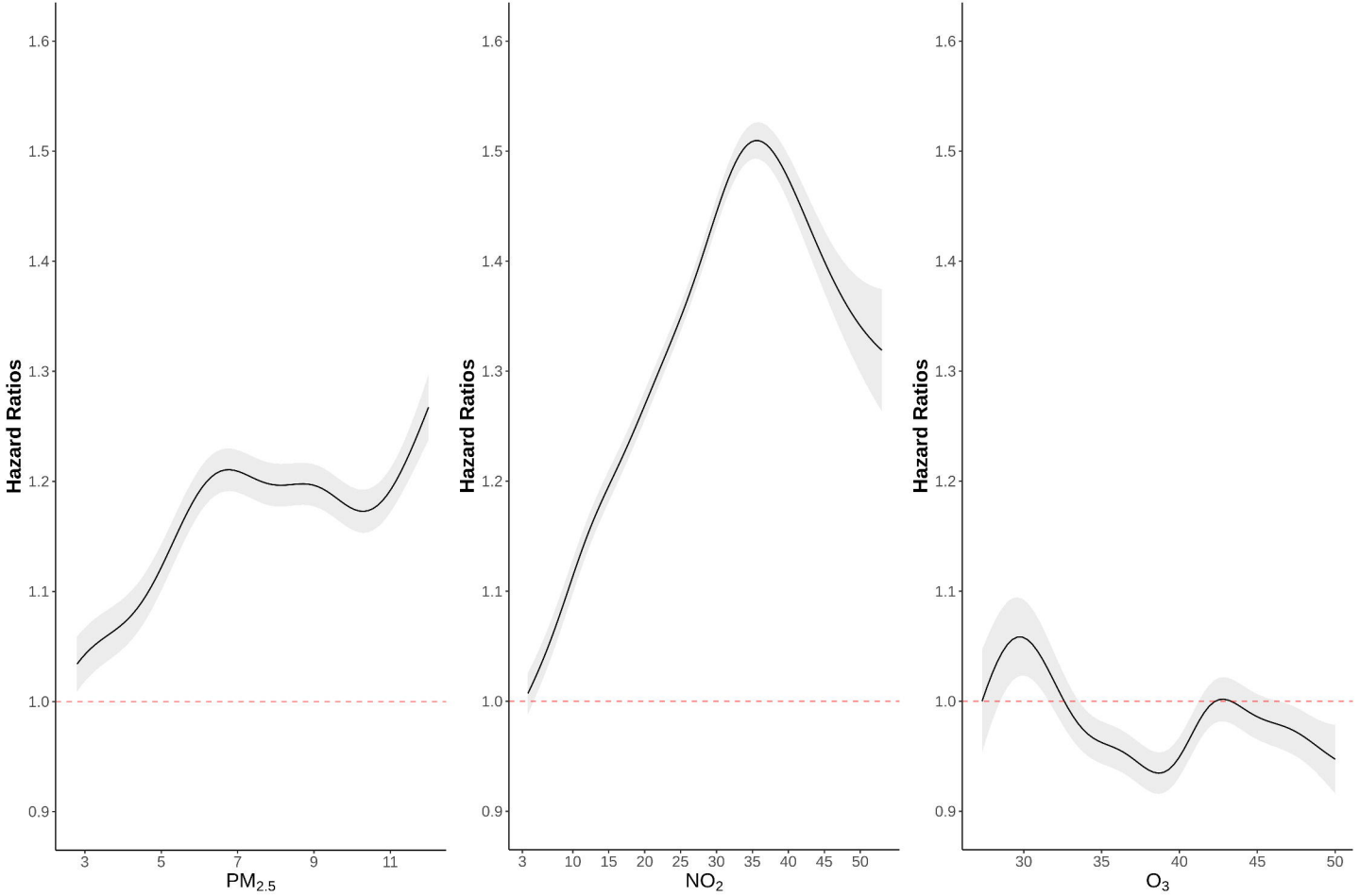
Figure 1. The association between $PM_{2.5}$, NO_2, and warm-months O_3 exposure and diabetes occurrence.

Figure 1 shows the exposure-response curves for the association between $PM_{2.5}$, NO_2, and warm-months O_3 and diabetes occurrence. We show the curve from the 0.5^{th} percentile of $PM_{2.5}$, i.e., with 0.5% poorly constrained extreme values excluded. Results are obtained from a multivariate model, adjusted for age, race, sex, Medicaid insurance, annual ZIP code means of summer and winter temperature, and annual ZIP code level sociodemographic variables. The model also includes a random intercept for each ZIP code and a spline function of year.

Figure 2. The association between low-levels air pollutants exposure and diabetes occurrence.

Figure 2 shows the exposure-response curves for the association between $PM_{2.5}$, NO_2, and warm-months O_3 and diabetes occurrence, in subsets, restricted to ZIP codes in which the annual exposures levels have not exceeded the national ambient air quality standards during the study period (i.e., $PM_{2.5}$<12 μg/m^3, NO_2<35 ppb, or O_3<50 ppb). We show the curves from the 0.5^{th} percentile of NO_2, i.e., with 0.5% poorly constrained extreme values excluded. Results are obtained from a multivariate model, adjusted for age, race, sex, Medicaid insurance, annual ZIP code means of summer and winter temperature, and annual ZIP code level sociodemographic variables. The model also includes a random intercept for each ZIP code and a spline function of year.


Hazard Ratios
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
Frequency
1.5e+07
1.0e+07
5.0e+06
0.0e+00
5
7
9
11
13
15
17
19
PM2.5
Hazard Ratios
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
Frequency
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
5
10
15
20
25
30
35
40
45
50
NO2
Hazard Ratios
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
Frequency
2e+07
1e+07
0e+00
30
35
40
45
50
55
60
O3


Hazard Ratios
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
3
5
7
9
11
PM2.5
Hazard Ratios
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
3
10
15
20
25
30
35
40
45
50
NO2
Hazard Ratios
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
30
35
40
45
50
O3