Evaluating the Joint Effects of O3, PM2.5, and NO2 as an Exposure Mixture on Mortality

Student: Nisha Lingam | Practicum Supervisor: Dr. Xiao Wu | Practicum Organization: N/A

I. Overview and Student Role

Long-term exposure to O_3, $PM_{2.5}$, and NO_2 has been consistently associated with all-cause mortality; however, the mixture effects of these pollutants are lesser-known.[1] The project aim is to investigate the overall effect of long-term exposure to $PM_{2.5}$, NO_2, and O_3 as a mixture of exposure species on all-cause mortality at the county-level in New York State from 2000-2016.

Mortality was measured as the total count of deaths in each county per month. Daily exposure concentrations for 1km x 1km grid cells were acquired using high-resolution, predictive models, which were aggregated by county and averaged by month.[2] Covariates included demographic, environmental, and socioeconomic variables. To estimate the exposure mixture effects, two approaches were used: weighted quantile sum (WQS) regression and quantile g-computation. WQS regression involves categorically transforming the exposures defined by quantiles and creating an exposure index from the weighted average of the transformed exposures, which is then used in a generalized linear model. Quantile g-computation is an extension of WQS regression that reduces bias and can apply directional heterogeneity to resulting regression models and the effect estimates. The student performed the data processing and analysis for this project, aided by data sourcing and suggestions from the practicum supervisor.

II. Background

Most literature shows that exposure to air pollution is associated with higher mortality rates. Specifically, $PM_{2.5}$ (particulate matter less than 2.5 μM in diameter) and NO_2 and their associations with all-cause mortality are well-documented, with both short-term and long-term exposure to these pollutants found to be associated with increases in all-cause mortality.[3] Additionally, increasing ambient O_3 concentrations produced by "atmospheric chemical reactions" between compounds such as nitrogen oxides and precursor gas emission VOC's (volatile organic compounds) is an emergent cause of health loss and death, which contribute to the overall risk factors for the worldwide disease burden.[4] The combined effects of pollutants such as O_3, $PM_{2.5}$, and NO_2 make up a smaller portion of the available literature and are accordingly, less understood. The WHO notes that their current pollutant guidelines provide recommendations for individual air pollutants and do not take into account these combined or mixture effects due to available evidence focusing on single-pollutant models to evaluate the effects of air pollution on human health.[5]

Studies that aim to characterize these combined effects use various statistical analyses, including generalized linear regressions with the inclusion of random effects, Cox-proportional hazards models, shrinkage methods, environment-wide association studies (EWAS), Bayesian kernel machine regression (BKMR), and weighted quantile sum (WQS) regressions. WQS regression and quantile generalized-computation (g-computation) are both methods that have been used to determine multi-pollutant mixture health effects and "rank" these major mixture components, especially in studies interested in analyzing $PM_{2.5}$ as a mixture of components.[6, 7] For example, when analyzing several major $PM_{2.5}$ species as a mixture, long-term exposure to these individual components (from 2000-2018) have been found to be positively associated with all-cause mortality among Medicare patients.[8] Another study conducted using the Korea National Health and Nutrition Examination Survey (KNHANES) database investigated NO_2, O_3, $PM_{2.5}$, SO_2, and CO as mixture and its effects on small airway dysfunction (SAD) using quantile g-computation, finding that this combined effect was significantly associated with SAD.[9]

This study focuses on the mixture effects of NO_2, O_3, $PM_{2.5}$ on all-cause mortality in New York state on a county-level using WQS regression and quantile g-computation to quantify these effects and compare the two varying methodologies.

III. Data and Methods

The air pollution data was obtained from the daily exposure concentrations for 1km x 1km grid cells provided by NASA and SEDAC; these concentrations were acquired using high-resolution, predictive models, which were aggregated by county and averaged by month. NO_2 and O_3 was measured in ppb, while $PM_{2.5}$ was measured in $\mu g/m^3$.[2] All-cause mortality and population data was collected from the CDC's Wonder Online Database by county and month. Demographic and socioeconomic variables including race, median income, property-owner occupancy, median age were acquired from Dr. Xiao Wu's study on "Evaluating the impact of long-term exposure to find particulate matter on mortality among the elderly".[10] Normalized difference vegetation index (NDVI) was included as an environmental factor, which involved NASA's Moderate Resolution Imaging Spectroradiometer collecting real-time data over 16-day intervals.[11] NDVI was averaged by month at the county level. Poisson generalized linear models (GLMs) were built using each exposure component with a spatial (county-level) fixed effect to determine each pollutant's individual association with mortality. All data ranges from 2000-2016.

WQS regression transforms a set of exposures into a set of categorical variables that are defined by specified quantiles. Using a given outcome of interest and a set of other covariates, a regression model adjusted for the given covariates is produced between the outcome and index exposure. This method also produces a set of weights, estimated as the mean weights over bootstrapped samples - all of which can be positive or negative - that quantify the contribution of each individual exposure to the overall mixture effect estimate. Overfitting can be prevented and multicollinearity between exposure types can be overcome, allowing for the analysis of highly correlated exposures. Notably, WQS uses a constraint in the model to force all weights to be in the same direction and linear additivity is assumed, potentially resulting in biased effect estimates.[12]

$$Yi = \beta_0 + \psi S_i + \varepsilon_i \qquad \textit{Exposure index } S_i = \sum_{i=1}^{n} w_i * q_i$$

Quantile g-computation combines WQS and g-computation (typically defined in this scenario as the g-formula, or "the generalization of standardization that uses the law of total probability to compute estimates of the expected outcome distribution under specific exposure patterns").[12] After fitting the linear model, directional homogeneity is initially assumed and if it does not hold, the weights are redefined to be positive or negative. Nonlinearity can also be accounted for using g-computation through the addition of nonlinear terms. The key difference with quantile g-computation is the ability to assign each mixture component weight to be either positive or negative, as WQS requires that all components have weights in the same direction. In the case of uncertainty surrounding confounders, quantile g-computation is a quicker, more efficient method to determine the mixture effect of air pollutants on mortality.

R Statistical Software was used for all statistical analyses. Poisson generalized linear models were evaluated for NO_2, O_3, $PM_{2.5}$, county and month variables to ensure the relationships between these variables and the outcome follow a Poisson distribution. The gwqs()

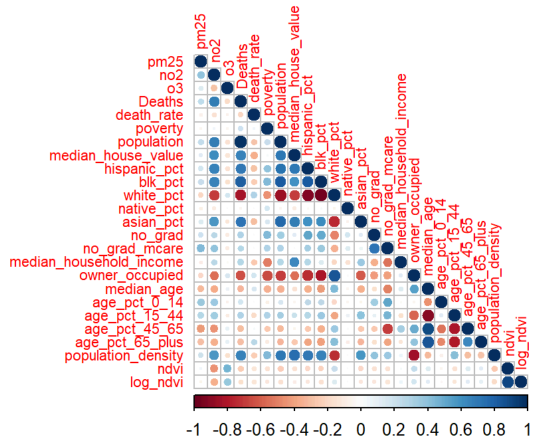
package was used to implement WQS regression with a quasi-Poisson distribution and in the positive direction to reflect the assumption that the exposures do not decrease mortality; total monthly mortality count on the county level was specified as the outcome. Similarly, the qgcomp() package was utilized with the same outcome and the Poisson distribution; however, a bidirectional association was allowed. All models also accounted for the following covariates: county-level proportions (ranging from 0 to 1) of white, Hispanic, black, Native American, and Asian demographic groups, county-level proportions of owner-occupied property status, median income, NDVI, and median age. Median income and NDVI were log-transformed and a log-transformed population offset were included in the final models. NDVI was additionally imputed using monthly data preceding or following the missing entries. Three models were produced using each method (six models total), with every model containing either a spatial fixed effect, a natural spline-adjusted temporal variable, or both factors. The spatial effect was fixed, due to the gwq() package and qgcomp() packages being unable to support the syntax for inclusion of random effects, as the lme4() package is required for random effect syntax, but is not compatible with the aforementioned packages. The spline-adjusted temporal variable was fitted with a df=10 based on models having the largest reduction in AIC values.

IV. Results

Table 1. Annual Median and Interquartile Ranges of Pollutant Concentrations

Year / Group	NO_2 Median	PM2.5 Median	O_3 Median	NO_2 IQR	PM2.5 IQR	O_3 IQR
Jan 2000	27.67137	11.398462	28.60369	[24.3, 29.9]	[9.7, 12.8]	[26.2, 29.5]
2000	24.27010	10.822509	37.44203	[19.2, 29.5]	[8.9, 12.5]	[29.7, 42.1]
2001	22.25143	10.656212	38.92905	[16.6, 28.9]	[9, 13.5]	[28.4, 48.7]
2002	21.70774	10.158488	38.08704	[15.7, 27.8]	[8.3, 13.1]	[29.2, 47.8]
2003	21.73400	10.480758	37.28849	[16, 28.5]	[8.6, 13.9]	[29.9, 45.2]
2004	18.64287	10.466333	37.63850	[13, 24.6]	[8.3, 12.7]	[28.5, 43.3]
2005	17.14369	10.862458	40.26062	[11.7, 25.6]	[8, 14.1]	[28.7, 46.2]
2006	16.92134	9.024298	36.75163	[11.8, 22.9]	[7.4, 10.6]	[28.4, 43.1]
2007	14.50647	9.749826	38.52025	[10.4, 20.9]	[7.9, 11.9]	[29.7, 43.3]
2008	14.19513	8.064566	35.30332	[10, 19.7]	[6.9, 10.6]	[29.3, 44.8]
2009	13.15856	7.652077	36.88171	[8.9, 19.6]	[6.2, 9.3]	[30.4, 41.7]
2010	12.97507	7.385521	40.18876	[9.1, 17.8]	[6.2, 9.1]	[33.4, 45.3]
2011	14.12242	7.890413	38.98913	[10.6, 20.5]	[6.4, 9.9]	[33.7, 42.3]
2012	15.10442	7.658134	38.48742	[11.8, 19]	[6.5, 9]	[32.7, 44.4]
2013	13.56130	7.554875	39.40759	[10.6, 17.9]	[6.1, 9.3]	[34.6, 43.6]
2014	13.78287	6.997437	39.59411	[9.9, 19.6]	[5.6, 8.2]	[34.1, 43]
2015	15.80019	7.140720	39.51418	[12.9, 21.8]	[5.5, 8.5]	[34.6, 43.1]
2016	13.95598	5.760535	39.52854	[11.7, 17.6]	[4.9, 6.7]	[34.1, 44.3]

Fig. 1. Correlation Matrix for Exposures, Covariates, and Outcome



The correlations between the exposures were relatively weak, while correlations between mortality and certain covariates such as race, owner occupied property, and median house value were fairly strong (> 0.6).

Table 2. Bootstrapped Poisson Generalized Linear Models for Exposures (adjusted for Spatial Fixed Effect)

Exposure	Coefficient	Rate Ratio	2.5% CI	97.5% CI
O_3	-0.00433	0.9957	0.9955	0.9959
$PM_{2.5}$	0.00248	1.0025	0.9980	1.0058
NO_2	0.00419	1.0042	1.0027	1.0055

When adjusted for county fixed effects, O_3 has a slight negative association with mortality, while $PM_{2.5}$ has no significant association with mortality, and NO_2 has a slight positive association with mortality.

Table 3. gWQS Model 1 - Spatial & Temporal Effects Included

	Rate Ratio	Lower 95% CI	Upper 95% CI
wqs	1.00569	1.00380	1.00759
pm25	0.99932	0.99859	1.00006
no2	1.00159	1.00132	1.00186
o3	0.99631	0.99605	0.99657
white_pct	0.84588	0.71226	1.00455
hispanic_pct	0.55140	0.45718	0.66504
blk_pct	8.91013	6.53855	12.14191
native_pct	0.09244	0.03381	0.25271
asian_pct	0.78514	0.61021	1.01020
log_med_income	0.93456	0.87458	0.99865
owner_occupied	0.67247	0.55738	0.81132
median_age	1.03588	1.03352	1.03825
log_ndvi	0.96909	0.96670	0.97149

When accounting for spatial and temporal confounding, a decile increase in the levels of the three exposure species is associated with an increase in the expected all-cause mortality rate by a factor of 1.006, or 0.6% (95% CI: [1.004, 1.008]), conditional on all other covariates. Exposures included as covariates represent the independent effects of each pollutant after accounting for the mixture effect. The proportion of black residents has a significant positive association with mortality count - a 10% increase in this proportion is associated with an expected all-cause mortality rate increase by a factor of 1.2445, or 24.45%.

Table 4. gWQS Model 2 - Spatial Fixed Effect Only

	Rate Ratio	Lower 95% CI	Upper 95% CI
wqs	1.00404	1.00199	1.00610
pm25	0.99935	0.99863	1.00006
no2	1.00162	1.00133	1.00190
o3	0.99662	0.99633	0.99692

When accounting for spatial confounding alone, a decile increase in the levels of the three exposure species is associated with an increase in the expected all-cause mortality rate by a factor of 1.004, or 0.4% (95% CI: [1.002, 1.006]), conditional on all other covariates. Exposures included as covariates represent the independent effects of each pollutant after accounting for the mixture effect. Remaining covariates were excluded from the results since their associations resemble those of Model 1 from Table 2.

Table 5. gWQS Model 3 - Spline-Adjusted Temporal Effect Only

	Rate Ratio	Lower 95% CI	Upper 95% CI
wqs	1.02594	1.02388	1.02801
pm25	0.99960	0.99863	1.00058
no2	1.00129	1.00093	1.00165
o3	0.99418	0.99381	0.99454

When accounting for temporal confounding alone, a decile increase in the levels of the three exposure species is associated with an increase in the expected all-cause mortality rate by a factor of 1.026, or 0.26% (95% CI: [1.024, 1.028]), conditional on all other covariates. Exposures included as covariates represent the independent effects of each pollutant after accounting for the mixture effect. Remaining covariates were excluded from the results since their associations resemble those of Model 1 from Table 2.

Tables 6, 7, and 8. Model Exposure Weights [Estimates & 95% CIs]

T5. gWQS Model 1

Pollutant	Estimate	Lower CI	Upper CI
NO_2	0.34730	0.33991	0.36340
O_3	0.33959	0.33304	0.34966
$PM_{2.5}$	0.31310	0.28869	0.32285

T6. gWQS Model 2

Pollutant	Estimate	Lower CI	Upper CI
NO_2	0.34572	0.33916	0.36846
O_3	0.33825	0.33327	0.34729
$PM_{2.5}$	0.31603	0.28416	0.32588

T7. gWQS Model 3

Pollutant	Estimate	Lower CI	Upper CI
NO_2	0.52513	0.48179	0.57609
O_3	0.35472	0.32323	0.38129
$PM_{2.5}$	0.12015	0.08143	0.16085

For the models accounting for spatial confounding, contributions to the overall mixture effect were similarly sized among all three pollutants. However, in the model accounting for temporal confounding alone, NO_2 has the largest contribution, while $PM_{2.5}$ has the smallest contribution, and O_3 has a similar contribution to those of Models 1 and 2.

Table 9. QG Comp 3-Model Summary

Model	Effect Estimate	Lower CI	Upper CI	Rate Ratio	RR Lower CI	RR Upper CI	p-value
Spatial Fixed Effect and Temporal Spline-Adjusted Variable	1.3477e-03	-3.6202e-04	3.0573e-03	1.0013	0.99964	1.0031	0.12236
Spatial Fixed Effect	1.7966e-03	2.0635e-04	3.3868e-03	1.0018	1.0002	1.0034	0.026809
Temporal Spline-Adjusted Variable	1.0469e-02	8.915e-03	1.2022e-02	1.0105	1.009	1.0121	0

Three separate models derived using QGComp are summarized with their overall mixture effect estimates, confidence intervals, and p-values. The overall mixture effect from quantile g-computation is interpreted as the effect on the outcome of increasing every exposure by one quantile, conditional on covariates. For the model accounting for both spatial and temporal confounding factors, each decile increase in the levels of the three exposure species as mixture is not associated with a significant change in the expected all-cause mortality rate (p-value > α=0.05). For the model accounting for spatial confounding alone, with each decile increase in the pollutant mixture, the expected all-cause mortality rate increases by a factor of 1.0018, or by 0.18% (95% CI: [1.0002, 1.0034]), conditional on all other covariates. For the model accounting for temporal confounding alone, with each decile increase in the levels of the three exposure species, the expected all-cause mortality rate increases by a factor of 1.0105 or by 0.105% (95% CI: [1.009, 1.012]), conditional on all other covariates.

V. Discussion/Conclusions

Overall, the results indicate that NO_2, O_3, and $PM_{2.5}$ may not have a large mixture effect associated with mortality. The quantile g-computed models suggest insignificant (p-value < α=0.05) associations when accounting for both spatial and temporal confounding. When ceasing to control for temporal confounding, the positive association is slightly stronger, relative to controlling for both factors (RR: 1.0018). The qgcomp model accounting for temporal confounding alone showed an even stronger positive association between the pollutant mixture and mortality at a rate ratio of 1.0105. In contrast, the WQS model showed a slightly stronger positive association in the model accounting for spatial and temporal confounding (RR: 1.0060), compared to the model accounting for spatial confounding alone (RR: 1.0040). However, similar to the qgcomp models, the model adjusting for temporal confounding alone showed the largest positive association (RR: 1.0260). The pollutant weights for the WQS models experience major changes with the temporal confounding-only model as well. These differences suggest that spatial and county-level factors may play a role in the mixture effect which requires further adjustment and study. The qgcomp models did not have the pollutant weights available to report, due to the bootstrap function being utilized to produce these models. Otherwise, the mixture effect estimates resulting from both WQS regression and quantile g-computation were relatively similar.

A major strength of this study is the use of two mixture analysis methods, considering that quantile g-computation ensures a reduction in bias that may have been present in the WQS regression results. These methods avoid overfitting and collinearity while determining unbiased mixture effects of NO_2, O_3, and $PM_{2.5}$. All models adjusted for many different demographic and socioeconomic variables, and the consideration of spatial and temporal confounding factors further reduced bias in the resulting estimates. There are some limitations with this study. The dataset was restricted to New York state counties, inevitably leading to a lower statistical power due to the smaller sample size and decreased variability. Moreover, there were difficulties in conducting subgroup analyses especially, since the data is on a county-population level, rather than an individual level. A couple noted weaknesses of WQS and quantile g-computation are the loss of information when abstracting the pollutant concentrations into quantiles and the reduced statistical power linked to splitting the data into training and validation sets.[13]

Additionally, the use of a fixed effect to account for spatial confounding rather than a random effect may not have accounted for the true variability associated with the mixture effect. This study was also restricted to a single outcome; however, it may have been useful or interesting to consider cause-specific mortality (i.e. cardiovascular or respiratory-related mortality) or non-accidental mortality alone. Mortality related to respiratory or cardiovascular causes could be an additional relevant outcome, considering the evidence available which supports a causal relationship between atmospheric O_3 levels and respiratory illnesses.[14] The inclusion of temperature or other related meteorological data could have provided opportunities to study the mixture effect of NO_2, O_3, $PM_{2.5}$ and any potential differences that may occur in different climates. It could also be worthwhile to further divide $PM_{2.5}$ into its major components to obtain a better understanding of any potential mixture effects between specific $PM_{2.5}$ components. Other studies have also utilized different quantiles - for example the KHANES study used quartiles, which may better account for the differences in measurement units and distributions for the individual pollutants. In conclusion, when considering the mixture effect of NO_2, O_3, and $PM_{2.5}$ on all-cause mortality, there are no significant, large associations. Further study is warranted using mixture analyses and a wider geographic range of data.

VI. References

1. Turner MC, Jerrett M, Pope CA 3rd, Krewski D, Gapstur SM, Diver WR, Beckerman BS, Marshall JD, Su J, Crouse DL, Burnett RT. Long-Term Ozone Exposure and Mortality in a Large Prospective Study. Am J Respir Crit Care Med. 2016 May 15;193(10):1134-42. doi: 10.1164/rccm.201508-1633OC. PMID: 26680605; PMCID: PMC4872664.
2. Yaguang Wei1, Xiaoshi Xing2, Alexandra Shtein1, Edgar Castro1, Carolynne Hultquist2,3, Mahdieh Danesh Yazdi1,4, Longxiang Li1, and Joel Schwartz1. 2022. Daily and Annual PM2.5, O3, and NO2 Concentrations at ZIP Codes for the Contiguous U.S., 2000-2016, v1.0. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/9yp5-hz11.
    2.1. Harvard T.H. Chan School of Public Health, Boston, MA, United States
    2.2. The Center for International Earth Science Information Network (CIESIN), Columbia University, Palisades, NY, United States
    2.3. School of Earth and Environment, University of Canterbury, Christchurch, New Zealand
    2.4. Stony Brook University, New York, United States
3. Pablo Orellano, Julieta Reynoso, Nancy Quaranta, Ariel Bardach, Agustin Ciapponi, Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and meta-analysis, Environment International, Volume 142, 2020, 105876, ISSN 0160-4120, https://doi.org/10.1016/j.envint.2020.105876.
4. Wan Hu, Junnan Yang, Effect of ambient ozone pollution on disease burden globally: A systematic analysis for the global burden of disease study 2019, Science of The Total Environment, Volume 926, 2024, 171739, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2024.171739.
5. World Health Organization. (2021). WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization. https://iris.who.int/handle/10665/345329. License: CC BY-NC-SA 3.0 IGO

6. Jin, T., Amini, H., Kosheleva, A. *et al.* Associations between long-term exposures to airborne PM2.5 components and mortality in Massachusetts: mixture analysis exploration. *Environ Health* 21, 96 (2022). https://doi.org/10.1186/s12940-022-00907-2
7. Mahdieh Danesh Yazdi, Heresh Amini, Yaguang Wei, Edgar Castro, Liuhua Shi, Joel D. Schwartz, Long-term exposure to PM2.5 species and all-cause mortality among Medicare patients using mixtures analyses, Environmental Research, Volume 246, 2024, 118175, ISSN 0013-9351, https://doi.org/10.1016/j.envres.2024.118175.
8. H. Boogaard, A.P. Patton, R.W. Atkinson, J.R. Brook, H.H. Chang, D.L. Crouse, J.C. Fussell, G. Hoek, B. Hoffmann, R. Kappeler, M. Kutlar Joss, M. Ondras, S.K. Sagiv, E. Samoli, R. Shaikh, A. Smargiassi, A.A. Szpiro, E.D.S. Van Vliet, D. Vienneau, J. Weuve, F.W. Lurmann, F. Forastiere, Long-term exposure to traffic-related air pollution and selected health outcomes: A systematic review and meta-analysis, Environment International, Volume 164, 2022, 107262, ISSN 0160-4120, https://doi.org/10.1016/j.envint.2022.107262.
9. Lee HW, Lee HJ, Oh S, Lee JK, Heo EY, Kim DK. Combined effect of changes in NO2, O3, PM2.5, SO2 and CO concentrations on small airway dysfunction. Respirology. 2024 May;29(5):379-386. doi: 10.1111/resp.14687. Epub 2024 Feb 20. PMID: 38378265.
10. Wu X, Braun D, Schwartz J, Kioumourtzoglou MA, Dominici F. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. Sci Adv. 2020 Jul 17;6(29):eaba5692. doi: 10.1126/sciadv.aba5692. PMID: 32832626; PMCID: PMC7439614.
11. Kamel Didan - University of Arizona, Alfredo Huete - University of Technology Sydney and MODAPS SIPS - NASA. (2015). MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid. NASA LP DAAC. http://doi.org/10.5067/MODIS/MOD13Q1.006
12. Keil AP, Buckley JP, O'Brien KM, Ferguson KK, Zhao S, White AJ. A Quantile-Based g-Computation Approach to Addressing the Effects of Exposure Mixtures. Environ Health Perspect. 2020 Apr;128(4):47004. doi: 10.1289/EHP5838. Epub 2020 Apr 7. PMID: 32255670; PMCID: PMC7228100.
13. Yu L, Liu W, Wang X, Ye Z, Tan Q, Qiu W, Nie X, Li M, Wang B, Chen W. A review of practical statistical methods used in epidemiological studies to estimate the health effects

of multi-pollutant mixture. Environ Pollut. 2022 Aug 1;306:119356. doi: 10.1016/j.envpol.2022.119356. Epub 2022 Apr 27. PMID: 35487468.

14. U.S. EPA. Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants (Final Report, Apr 2020). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-20/012, 2020.