Supplemental Information for:

Residential proximity to biorefinery sources of air pollution and respiratory diseases in New York State

Eun Kyung Lee[a,b], Xiaobo Xue Romeiko[a], Wang-Jian Zhang[a], Beth J. Feingold[a], Haider A. Khwaja[a,c], Xuesong Zhang[d,e], Shao Lin[a,f*]

[a] Department of Environmental Health Sciences, University at Albany, State University of New York, School of Public Health, 1 University Place, Rensselaer, New York 12144, USA

[b] Mary Ann Swetland Center for Environmental Health, Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, 11000 Cedar Avenue, Cleveland, Ohio 44106, USA

[c] Wadsworth Center, New York State Department of Health, Empire State Plaza, Albany, New York 12201, USA

[d] Joint Global Change Research Institute, Pacific Northwest National Laboratory, 5825 University Research Court, College Park, Maryland 20740, USA

[e] Earth System Sciences Interdisciplinary Center, University of Maryland, 5825 University Research Court, College Park, Maryland 20740, USA

[f] Department of Epidemiology and Biostatistics, University at Albany, State University of New York, School of Public Health, 1 University Place, Rensselaer, New York 12144, USA

***Corresponding author:**

Shao Lin, MD, PhD, MPH
Professor
Department of Environmental Health Sciences
School of Public Health, University at Albany
1 University Place, Room 212D
Rensselaer, NY 12144
Email: slin@albany.edu

This file includes:

Figures: 6

Tables: 6

Pages: 14

FIGURES

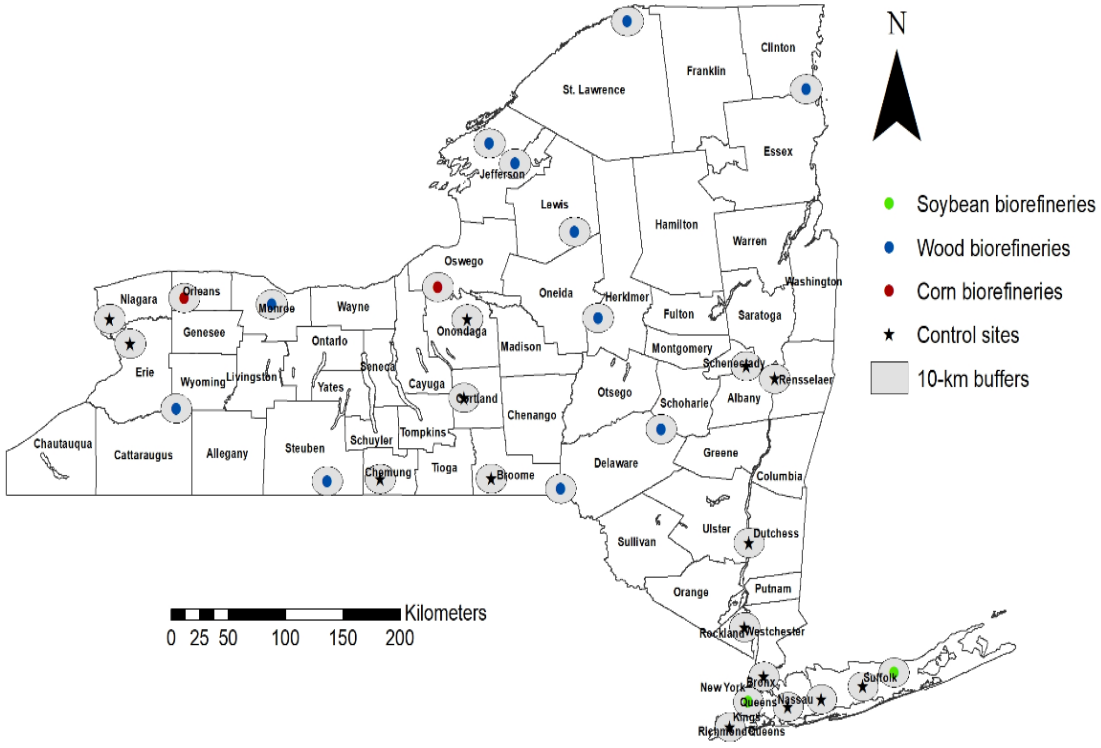

Figure S1. Study sites showing 15 biorefinery locations (dots), 15 control sites (stars) and 10 km buffers (grey circles) in New York State.

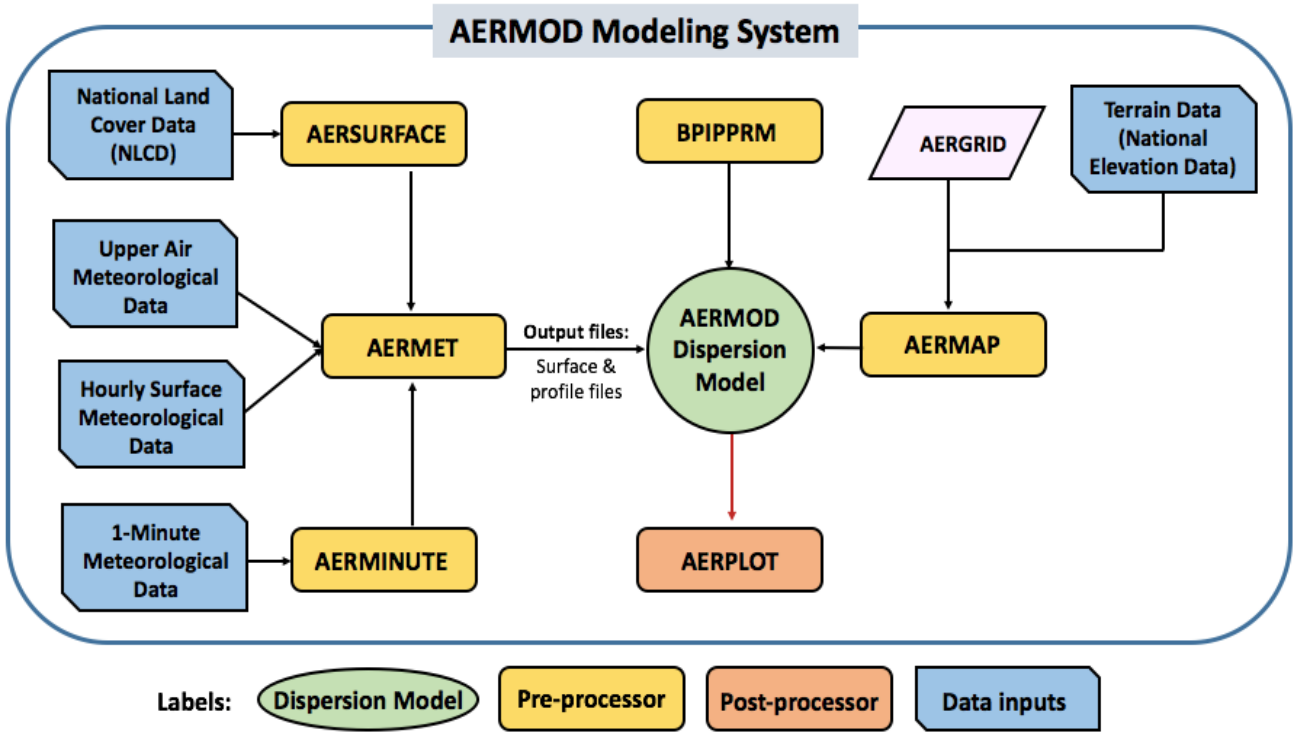

Figure S2. Diagram of AERMOD air dispersion model system.

Figure adapted from: http://tools.envirolink.govt.nz/dsss/aermod/

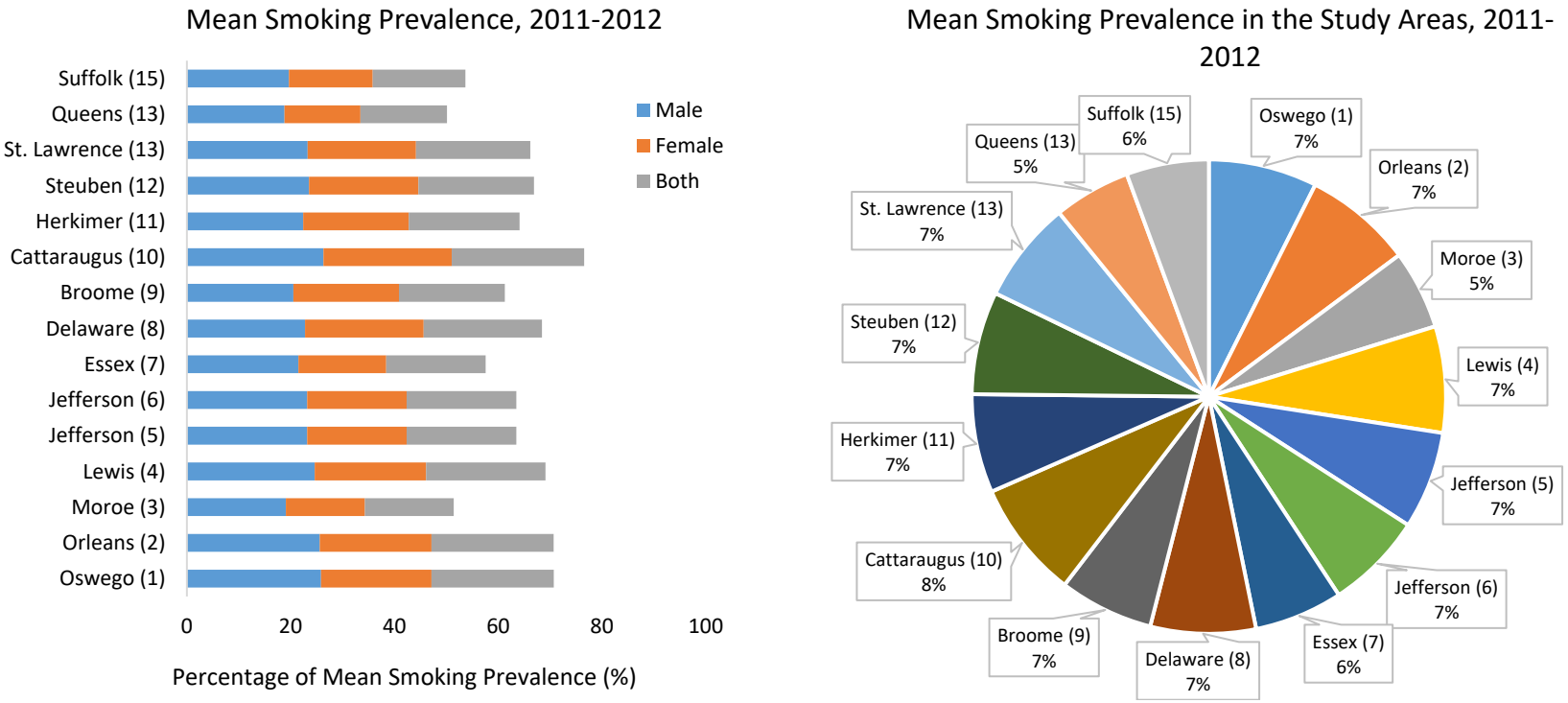

Figure S3. Mean smoking prevalence in study areas by county within New York State, 2011-2012.[1]

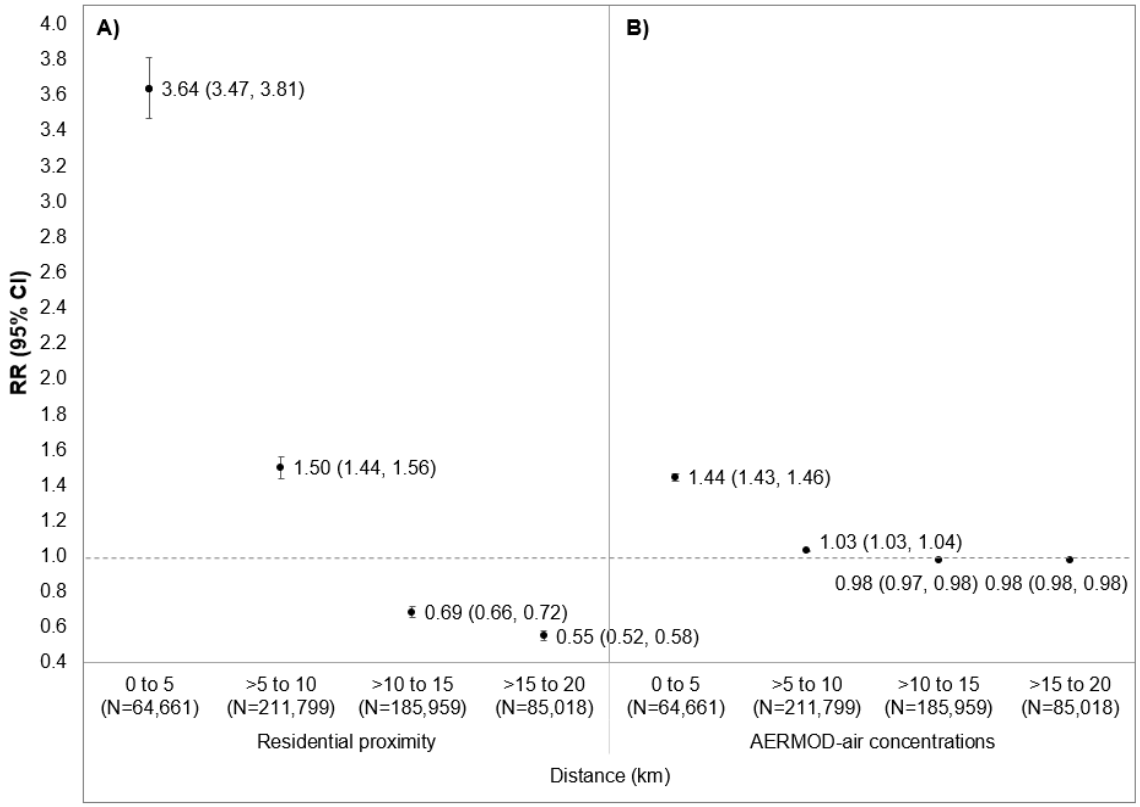

Figure S4. Adjusted rate ratios (RRs) and 95% confidence intervals of the associations between respiratory diseases with: A) residential proximity to biorefineries, and B) AERMOD-modeled air concentrations up to 20 km from 15 biorefinery facilities ("biorefinery sites") compared to matched areas ("control areas") in New York State, 2011-2015. All models were adjusted for age, race, sex, county-level smoking rate, seasonal mean temperature and relative humidity, annual background air pollutants of $PM_{2.5}$, SO_2 and NO_2.

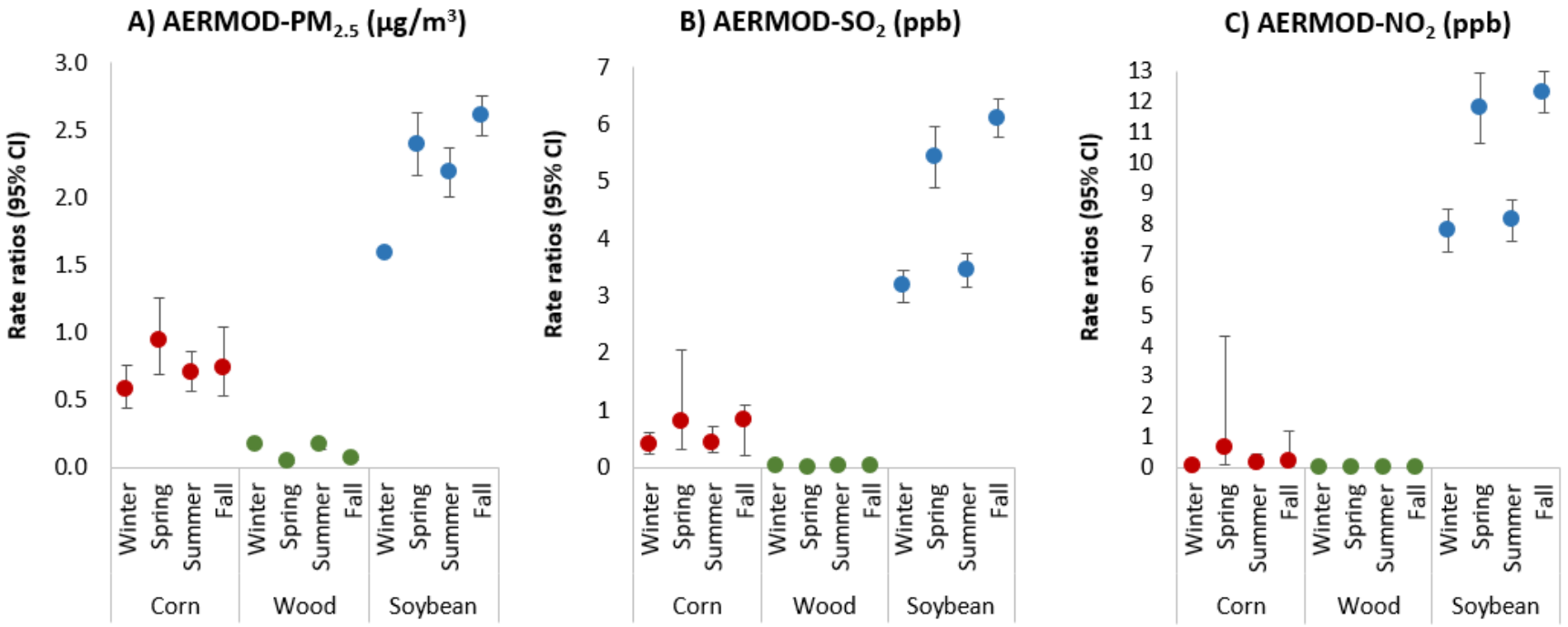

Figure S5. Adjusted rate ratios (RRs) and 95% confidence intervals of respiratory ED visits associated with each interquartile range change in AERMOD-modeled air concentrations of $PM_{2.5}$, SO_2 and NO_2 from 2011-2015 by seasons (winter, spring, summer and fall) and types of biorefineries (assessed within 10 km from the biorefineries). All models were adjusted for age, race, sex, county-level smoking rate, seasonal mean temperature and relative humidity, annual background air pollutants of $PM_{2.5}$, SO_2 and NO_2.

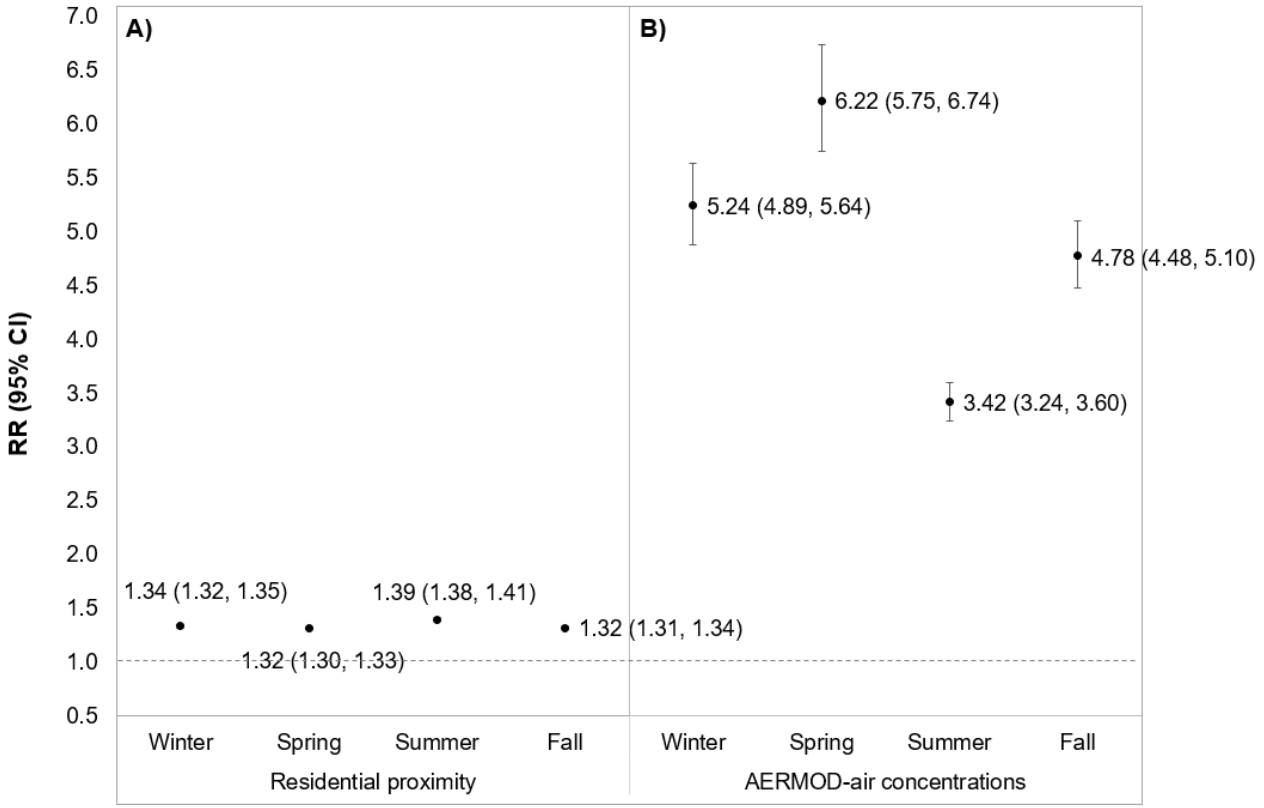

Figure S6. Adjusted rate ratios (RRs) and 95% confidence intervals of the associations between respiratory diseases and: A) residential proximity to biorefineries, and B) AERMOD-modeled air concentrations by seasons up to 10 km from 15 biorefinery facilities ("exposed areas") compared to control sites ("unexposed areas") in New York State, 2011-2015 (N=276,460). All models were adjusted for age, race, sex, county-level smoking rate, seasonal mean temperature and relative humidity, annual background air pollutants of $PM_{2.5}$, SO_2 and NO_2.

TABLES

Table S1. AERMOD air dispersion model parameters of the biorefinery facilities and stacks.

Bio-refinery No.	Type of biorefinery (feedstock)		Building [d]		Stacks [d]					City	Emission rate (g/s) [c]		
		Cap. MGY [d]	Ht (m)	Elev. (m)	Hgt. (m)	Elev. (m)	Temp. (K)	Vol. (m/s)	Dia. (m)		$PM_{2.5}$	SO_2	NO_2
1	Corn	85	9.3	116.0	34.9	116.0	738.7	25.0	3.0	Fulton	7.50	15.4	36.9
2	Corn	60	31.3	169.0	39.8	172.0	738.7	25.0	3.0	Medina	0.58	31.9	17.3
3	Wood	N/A	9.34	20.0	1.0 [a]	20.0 [b]	738.7	25.0	3.0	Rochester	0.33	7.48	2.47
4	Wood	22	14.7	333.7	24.0	331.6	738.7	25.0	3.0	Lyons Fall	7.50	15.4	36.9
5	Wood	60	20.0	50.0	24.0	50.0	738.7	25.0	3.0	Fort Drum	2.76	144.4	42.1
6	Wood	5.4	4.6	136.0	1.0 [a]	136.0 [b]	738.7	25.0	3.0	Keeseville	0.33	7.48	2.47
7	Wood	7.3	12.1	158.0	1.0 [a]	158.0 [b]	738.7	25.0	3.0	La Fargeville	0.33	7.48	2.47
8	Wood	N/A	12.1	20.0	1.0 [a]	20.0 [b]	738.7	25.0	3.0	Stamford	0.33	7.48	2.47
9	Wood	80	6.3	301.0	5.7	301.0	738.7	25.0	3.0	Deposit	0.33	7.48	2.47
10	Wood	90.7	8.9	449.0	3.0 [a]	449.0 [b]	738.7	25.0	3.0	Arcade	0.33	7.48	2.47
11	Wood	70	5.2	131.0	7.3	132.0	738.7	25.0	3.0	Frankfort	0.28	30.8	16.3
12	Wood	45	6.9	314.0	1.0 [a]	314.0 [b]	738.7	25.0	3.0	Addison	0.28	30.8	16.3
13	Wood	108.9	10.0	71.0	15.5	69.0	738.7	25.0	3.0	Massena	0.28	30.8	16.3
14	Soybean	50	22.4	13.0	15.0	13.0	738.7	25.0	3.0	Brooklyn	7.50	15.4	36.9
15	Soybean	N/A	10.9	16.0	1.0 [a]	16.0 [b]	738.7	25.0	3.0	Calverton	0.33	7.48	2.47

[a]Biorefineries with no stacks were assigned as 1.0 as a default; [b]Missing information of stacks' elevation were replaced by building's information; [c]Data source: State-level industrial emissions were obtained from the National Air Pollutant Emissions Inventory Trends data and literature-based values were used among biorefineries sharing similar building/stack characteristics.[2,3] For biorefineries with no information available, we replaced with information of biorefineries sharing similar characteristics; Abbreviations: MGY, million gallons yr^{-1}; Cap, capacity; Ht, height; Elev, elevation; Temp, temperature; Vol, volume; Dia, diameter; N/A: information not available

Average emission rates (g/s):

	Corn	Wood	Soybean
$PM_{2.5}$	4.0	1.19	3.9
SO_2	23.6	27.0	11.4
NO_2	27.1	12.9	19.7

Table S2. Data sources for simulating AERMOD air dispersion model.

Processor	Parameters	Data sources
AERSURFACE	National Land Cover data	National Land Cover Data (NLCD) (2011 data) from USGS: https://www.mrlc.gov/nlcd11_data.php
AERMINUTE	1-Minute Meteorological data (temperature/dew point, relative humidity, cloud cover, bowen ratio, windspeed, wind direction)	Automated Surface Observing Systems (ASOS): ftp://ftp.ncdc.noaa.gov/pub/data/asos-onemin/ National Climatic Data Center (NCDC) NOAA: http://raob.fsl.noaa.gov/
AERMET	Upper Air Meteorological data Hourly Surface Meteorological data	National Weather Service (NWS): https://www.weather.gov/ National Oceanic and Atmospheric Administration (NOAA): http://raob.fsl.noaa.gov/
AERMAP	Terrain data (National Elevation Data)	USGS MRLC: https://mrlc.gov/viewerjs/
BPIPPRM	Building information (building height, length, stack height, length, diameter, UTM coordinates, elevation, stack diameter (m), stack temperature (K), stack velocity (m/s)) Air pollutant emissions rate (g/s)	Google Earth Pro National Emissions Inventory

Steps for processing AERMOD model:

The AERMOD model is composed of three pre-processors: (1) the AERMOD Meteorological Preprocessor (AERMET), (2) the AERMOD Terrain Preprocessor (AERMAP), and, (3) the Building Profile Input Program for PRIME (BPIPPRM) (See Figure S1 and Table S1 for detailed information about the pre-processors).[4,5] For processing AERMET (v.18081), we obtained 1-minute Automated Surface Observing System (ASOS) wind data, hourly surface air data and upper air data from 2011-2015. The meteorological data were based on the closest weather station from each biorefinery and were obtained from the National Climatic Data Center (NCDC) and the National Oceanic and Atmospheric Administration (NOAA)/Earth System Research Laboratory (ESRL) radiosonde database.[6] To account for the surface characteristics, the land cover information was processed based on the National Land Cover Data (NLCD).[7] For processing AERMAP (v.18081), which determines the elevation and hill height scale for each of the buildings, the National Elevation Data (NED) was used. For processing BPIPPRM, a pre-processor to mimic building profiles and air stack downwash values, we included the amount of state-level industrial emissions data (emission rate of pollutants (g/s)) and stack-related data (i.e., stack height, temperature, diameter, velocity) for each biorefinery. We retrieved the state-level industrial emissions data from the National Air Pollutant Emissions Inventory Trends database and literature-based values (see Table S3).[2,3] The geographic locations and characteristics of the buildings (e.g., UTM coordinates of the buildings, building's height and wide) were retrieved from Google Earth.

Table S3. Statistical distribution of AERMOD-modeled air concentrations of $PM_{2.5}$, SO_2 and NO_2 by proximity distance (0-20 km) to biorefineries in New York State.

Distances from the biorefineries	Mean	Min	Max	Percentiles			
				25th	50th	75th	IQR
$PM_{2.5}$ ($\mu g/m^3$)							
0-5 km	0.564	0	11.2	0	0.001	0.077	0.077
>5-10 km	0.070	0	1.15	0	0.001	0.019	0.019
>10-15 km	0.066	0	0.76	0	0.001	0.015	0.015
>15-20 km	0.031	0	0.47	0	0.001	0.008	0.008
0-20 km	0.335	0	11.2	0	0.001	0.034	0.034
SO_2 (ppb)							
0-5 km	0.660	0	26.5	0.024	0.124	0.401	0.377
>5-10 km	0.080	0	0.48	0.014	0.045	0.104	0.091
>10-15 km	0.060	0	0.32	0.010	0.033	0.088	0.078
>15-20 km	0.040	0	0.23	0.006	0.024	0.054	0.047
0-20 km	0.390	0	26.5	0.013	0.066	0.204	0.190
NO_2 (ppb)							
0-5 km	1.109	0	55.5	0.013	0.099	0.413	0.400
>5-10 km	0.117	0	1.15	0.007	0.039	0.129	0.122
11-15 km	0.095	0	0.769	0.005	0.026	0.098	0.093
16-20 km	0.054	0	0.479	0.003	0.019	0.063	0.060
0-20 km	0.645	0	55.5	0.008	0.057	0.217	0.209

Table S4. Adjusted rate ratios (RRs) and 95% confidence intervals of the associations between respiratory morbidity and: A) residential proximity to biorefineries (up to 10 km), B) AERMOD-modeled air pollutants of $PM_{2.5}$, SO_2 and NO_2 covering up to 10 km from 15 biorefineries in New York State, 2011-2015.

	N (%)	Adjusted RR (95% CI)[d,e]	
		Asthma (N= 256,513)	**Other respiratory diseases**[c] (N= 19,947)
A. Residential proximity to biorefineries (0-10 km)[a]			
0-2 km	3,839 (1.39)	**3.51 (2.89, 4.28)**	**2.94 (1.58, 5.42)**
>2-4 km	32,228 (11.6)	**1.04 (1.03, 1.06)**	**1.00 (0.77, 1.30)**
>4-6 km	64,511 (23.3)	**1.45 (1.43, 1.47)**	**0.72 (0.69, 0.75)**
>6-8 km	77,296 (27.9)	**1.39 (1.38, 1.41)**	**2.34 (1.78, 3.06)**
>8-10 km	98,586 (35.7)	**2.09 (2.07, 2.11)**	**1.40 (1.36, 1.45)**
0-10 km	276,460 (100)	**1.45 (1.45, 1.46)**	**2.55 (1.08, 2.88)**
B. AERMOD-modeling air pollutants (0-10 km)[a,b]			
$PM_{2.5}$ ($\mu g/m^3$)			
0-2 km	3,839 (1.39)	**1.16 (1.13, 1.18)**	**1.13 (1.03, 1.23)**
>2-4 km	32,228 (11.6)	**1.24 (1.23, 1.26)**	**1.27 (1.19, 1.34)**
>4-6 km	64,511 (23.3)	**1.06 (1.06, 1.06)**	**1.04 (1.03, 1.06)**
>6-8 km	77,296 (27.9)	**1.05 (1.05, 1.05)**	**1.04 (1.03, 1.05)**
>8-10 km	98,586 (35.7)	**1.03 (1.03, 1.03)**	**1.03 (1.02, 1.03)**
0-10 km	276,460 (100)	**1.02 (0.98, 1.06)**	**1.08 (1.07, 1.09)**
SO_2 (ppb)			
0-2 km	3,839 (1.39)	**2.51 (2.21, 2.85)**	**2.13 (1.23, 3.69)**
>2-4 km	32,228 (11.6)	**1.86 (1.80, 1.92)**	**1.97 (1.66, 2.33)**
>4-6 km	64,511 (23.3)	**1.34 (1.32, 1.35)**	**1.25 (1.15, 1.35)**
>6-8 km	77,296 (27.9)	**1.22 (1.22, 1.23)**	**1.17 (1.12, 1.22)**
>8-10 km	98,586 (35.7)	**1.16 (1.15, 1.17)**	**1.14 (1.10, 1.17)**
0-10 km	276,460 (100)	**1.09 (0.88, 1.36)**	**1.52 (1.44, 1.61)**
NO_2 (ppb)			
0-2 km	3,839 (1.39)	**2.38 (2.12, 2.68)**	**2.04 (1.22, 3.43)**
>2-4 km	32,228 (11.6)	**2.28 (2.18, 2.39)**	**2.46 (1.97, 3.08)**
>4-6 km	64,511 (23.3)	**1.42 (1.40, 1.44)**	**1.31 (1.19, 1.44)**
>6-8 km	77,296 (27.9)	**1.32 (1.30, 1.33)**	**1.24 (1.17, 1.31)**
>8-10 km	98,586 (35.7)	**1.23 (1.22, 1.24)**	**1.20 (1.15, 1.25)**
0-10 km	276,460 (100)	**1.10 (0.87, 1.39)**	**1.57 (1.48, 1.66)**

[a]Adjusted by sex, age, race, county-level smoking, temperature, relative humidity and background air pollutants of $PM_{2.5}$, SO_2 and NO_2; [b]RRs for were estimated based on $e^{\beta*IQR}$; [c]Other respiratory diseases are: chronic bronchitis, emphysema and chronic airway osbstruction. These 3 subtypes were combined due to the small sample size, which caused overfitting in the model; [d]All values were statistically significant with p-values <0.05; [e]Abbreviations: RR, Rate ratio; CI, confidence intervals.

Table S5. Adjusted rate ratios (RRs) and 95% confidence intervals of the associations between respiratory diseases and: A) residential proximity (residency within 10 km) to 3 different types of biorefineries (corn, wood and soybean) and B) AERMOD-modeled air pollutants in New York State, 2011-2015.

Type of feedstock	N (%)	Adjusted RR (95% CI)		
		All respiratory (N=276,460)	**Asthma** (N= 256,513)	**Other respiratory diseases[c]** (N= 19,947)
A) Residential proximity to biorefineries (0-10 km)[a]				
All types	276,460 (100)	**1.67 (1.65, 1.69)**	**1.45 (1.45, 1.45)**	**2.55 (1.08, 2.88)**
Corn	1,332 (0.48)	**2.59 (2.53, 2.65)**	1.69 (0.04, 8.36)	**4.13 (3.88, 4.39)**
Wood	25,050 (9.06)	**1.66 (1.64, 1.69)**	**2.12 (1.98, 2.27)**	**1.54 (1.51, 1.58)**
Soybean	250,078 (90.5)	**1.83 (1.79, 1.87)**	**5.28 (3.95, 7.05)**	**4.43 (3.35, 5.87)**
B) AERMOD-modeling air pollutants (0-10 km)[a,b]				
Corn	1,332 (0.48)			
$PM_{2.5}$ (μg/m^3)		**2.048 (2.013, 2.083)**	1.488 (0.091, 4.944)	**2.906 (2.774, 3.045)**
SO_2 (ppb)		**1.352 (1.342, 1.362)**	1.182 (0.365, 1.959)	**1.567 (1.536, 1.598)**
NO_2 (ppb)		**1.922 (1.893, 1.952)**	1.436 (0.113, 4.292)	**2.644 (2.534, 2.759)**
Wood	25,050 (9.06)			
$PM_{2.5}$ (μg/m^3)		**1.003 (1.003, 1.003)**	**1.004 (1.003, 1.004)**	**1.002 (1.002, 1.002)**
SO_2 (ppb)		**1.072 (1.070, 1.074)**	**1.108 (1.098, 1.118)**	**1.061 (1.058, 1.064)**
NO_2 (ppb)		**1.056 (1.055, 1.058)**	**1.084 (1.076, 1.092)**	**1.048 (1.045, 1.050)**
Soybean	250,078 (90.5)			
$PM_{2.5}$ (μg/m^3)		**1.377 (1.361, 1.393)**	**2.410 (2.808, 2.067)**	**2.197 (2.549, 1.894)**
SO_2 (ppb)		**1.712 (1.679, 1.746)**	**4.391 (5.681, 3.394)**	**3.760 (4.827, 2.928)**
NO_2 (ppb)		**3.628 (3.462, 3.800)**	**34.6 (18.7, 64.2)**	**23.9 (13.1, 43.5)**

[a]Adjusted for sex, age, race, county-level smoking, temperature, relative humidity and background air pollutants; [b]RRs for were estimated based on $e^{\beta*IQR}$; [c] The 3 respiratory subtypes of chronic bronchitis, emphysema and chronic airway obstruction were combined due to the small number of cases for the 3 respiratory subtypes causing overfitting in the model; bolded values represent rate ratios with statistical significance of p-values <0.05.

Table S6. Sociodemographic and neighborhood characteristics of areas surrounding the 15 biorefineries.

Biorefinery No.	City	Urbanicity[a]	% Poverty	% Unemployed	Average per capita income	Lowest median income	Highest median income	% Disability	% Minority	% Uninsured
1	Fulton	Metro 1	11.6	14.7	$26,264.0	$26,264.0	$28,630.0	13.2	2.3	10.9
2	Medina	Metro 1	19.5	9.1	$22,578.0	$22,578.0	$25,879.0	14.1	16.3	15.7
3	Rochester	Metro 3	15.3	11.3	$18,495.0	$7,690.0	$42,683.0	15.6	23.7	7.4
4	Lyons Falls	Metro 3	21.5	15.7	$21,270.0	$21,270.0	$23,418.0	15.8	4	14
5	Fort Drum	Metro 1	17.9	13.7	$15,822.0	$19,915.0	$24,333.0	4.1	42.1	2.5
6	Keeseville	Metro 2	9.7	8.6	$28,787.0	$24,757.0	$28,787.0	15.5	4.3	8.7
7	La Fargeville	Metro 1	5.4	11.4	$25,046.0	$24,603.0	$31,412.0	12.1	19.3	7
8	Stamford	Metro 3	13.4	11	$24,674.0	$15,822.0	$31,711.0	14.9	4.7	7.1
9	Deposit	Metro 3	11	8.4	$23,139.0	$23,139.0	$25,479.0	14.5	5.9	11.6
10	Arcade	Metro 3	14.3	7.8	$23,122.0	$20,645.0	$26,386.0	16.4	2.8	5.6
11	Frankfort	Metro 3	9.1	9.7	$24,237.0	$21,237.0	$24,237.0	13	2.5	7.4
12	Addison	Metro 2	29.9	11.3	$19,519.0	$19,130.0	$19,519.0	14.7	3.2	20
13	Massena	Metro 2	11.3	15.9	$27,172.0	$19,852.0	$25,046.0	18.9	4.7	7.7
14	Brooklyn	Metro 1	18.2	3.6	$29,846.0	$12,864.0	$173,879.0	8.6	36	26.2
15	Calverton	Metro 1	6	12.8	$31,711.0	$31,035.0	$44,660.0	15.8	25	9.3
Average			**14.3**	**11.0**	**$24,112.1**	**$20,720.1**	**$38,403.9**	**13.8**	**13.1**	**10.7**
Median			**13.4**	**11.3**	**$24,237.0**	**$21,237.0**	**$26,386.0**	**14.7**	**4.7**	**8.7**
Min			**5.4**	**3.6**	**$15,822.0**	**$7,690.0**	**$19,519.0**	**4.1**	**2.3**	**2.5**
Max			**29.9**	**15.9**	**$31,711.0**	**$31,035.0**	**$173,879.0**	**18.9**	**42.1**	**26.2**

References

(1) GHDX. Global Health Data Exchange http://ghdx.healthdata.org/us-data (accessed Apr 1, 2020).

(2) Gibson, M. D.; Kundu, S.; Satish, M. Dispersion Model Evaluation of PM2.5, NOX and SO2 from Point and Major Line Sources in Nova Scotia, Canada Using AERMOD Gaussian Plume Air Dispersion Model. *Atmos. Pollut. Res.* **2013**, *4* (2), 157–167. https://doi.org/10.5094/APR.2013.016.

(3) US EPA. National air pollutant emissions inventory trends data https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data (accessed Mar 10, 2019).

(4) US EPA. AERMOD: Description of model formulation https://gaftp.epa.gov/Air/aqmg/SCRAM/models/preferred/aermod/aermod_mfed.pdf.

(5) US EPA. NAAQS table https://www.epa.gov/criteria-air-pollutants/naaqs-table (accessed Sep 8, 2019).

(6) NOAA. Automated surface observing system (ASOS) https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/automated-surface-observing-system-asos (accessed Feb 2, 2019).

(7) USGS. Multi-resolution land characteristics consortium https://www.mrlc.gov/tools (accessed Feb 10, 2019).