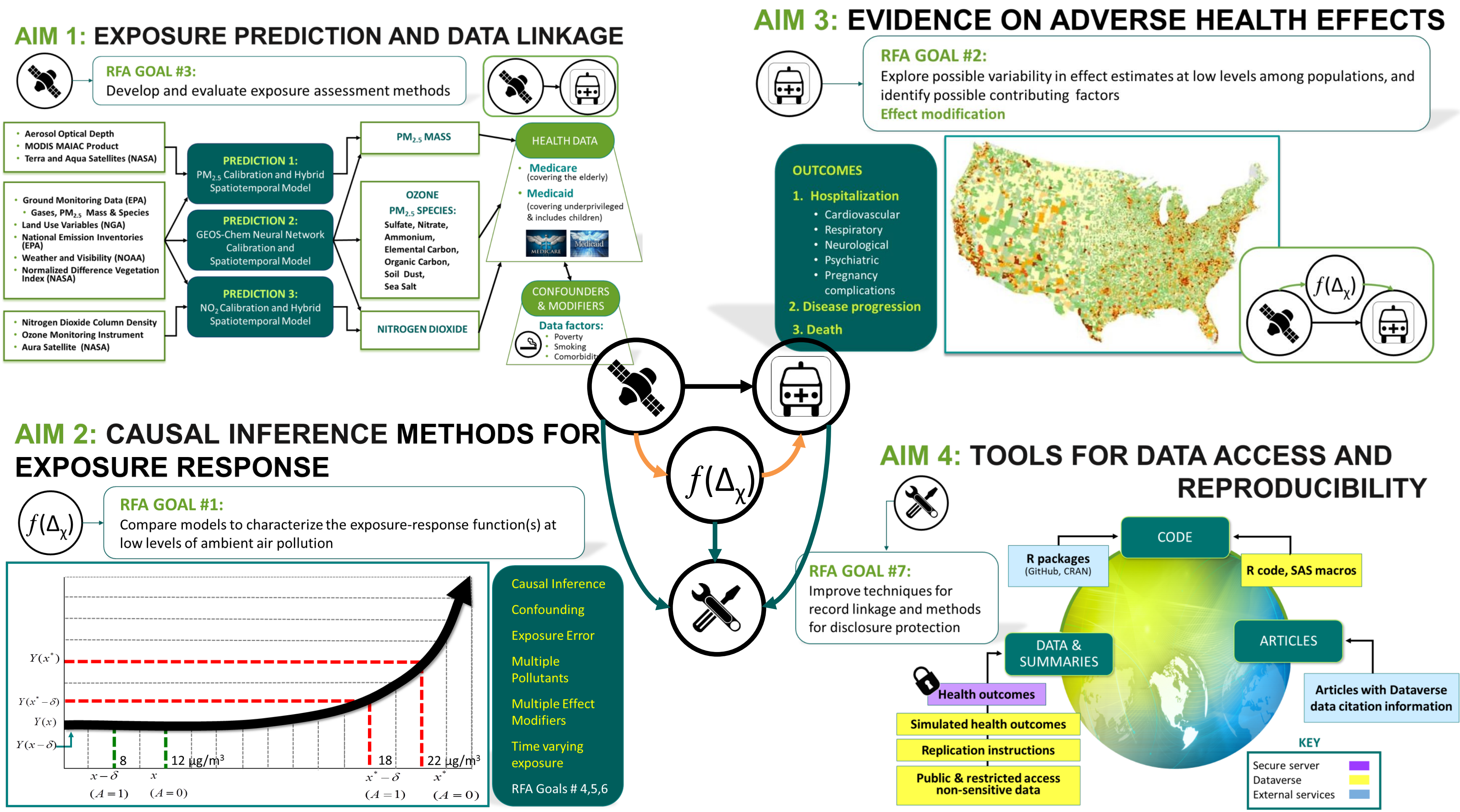
Results from the U.S. Study Using Medicare Data

Francesca Dominici, PhD
Clarence James Gamble Professor
of Biostatistics, Population and Data Science
Harvard T.H. Chan School of Public Health
Co-Director Harvard Data Science Initiative

HDSI | Harvard Data Science Initiative


AIM 1: EXPOSURE PREDICTION AND DATA LINKAGE
RFA GOAL #3:
Develop and evaluate exposure assessment methods
Aerosol Optical Depth
MODIS MAIAC Product
Terra and Aqua Satellites (NASA)
Ground Monitoring Data (EPA)
Gases, $PM_{2.5}$ Mass & Species
Land Use Variables (NGA)
National Emission Inventories (EPA)
Weather and Visibility (NOAA)
Normalized Difference Vegetation Index (NASA)
Nitrogen Dioxide Column Density
Ozone Monitoring Instrument
Aura Satellite (NASA)
PREDICTION 1:
$PM_{2.5}$ Calibration and Hybrid Spatiotemporal Model
PREDICTION 2:
GEOS-Chem Neural Network Calibration and Spatiotemporal Model
PREDICTION 3:
NO_2 Calibration and Hybrid Spatiotemporal Model
$PM_{2.5}$ MASS
OZONE
$PM_{2.5}$ SPECIES:
Sulfate, Nitrate, Ammonium, Elemental Carbon, Organic Carbon, Soil Dust, Sea Salt
NITROGEN DIOXIDE
HEALTH DATA
Medicare
(covering the elderly)
Medicaid
(covering underprivileged & includes children)
MEDICARE
Medicaid
CONFOUNDERS & MODIFIERS
Data factors:
Poverty
Smoking
Comorbidit
AIM 3: EVIDENCE ON ADVERSE HEALTH EFFECTS
RFA GOAL #2:
Explore possible variability in effect estimates at low levels among populations, and identify possible contributing factors
Effect modification
OUTCOMES
1. Hospitalization
Cardiovascular
Respiratory
Neurological
Psychiatric
Pregnancy complications
2. Disease progression
3. Death
$f(\Delta_\chi)$
AIM 2: CAUSAL INFERENCE METHODS FOR EXPOSURE RESPONSE
RFA GOAL #1:
Compare models to characterize the exposure-response function(s) at low levels of ambient air pollution
$Y(x^*)$
$Y(x^*-\delta)$
$Y(x)$
$Y(x-\delta)$
8
12 µg/m³
18
22 µg/m³
$x-\delta$
x
$x^*-\delta$
x^*
$(A=1)$
$(A=0)$
$(A=1)$
$(A=0)$
Causal Inference
Confounding
Exposure Error
Multiple Pollutants
Multiple Effect Modifiers
Time varying exposure
RFA Goals # 4,5,6
AIM 4: TOOLS FOR DATA ACCESS AND REPRODUCIBILITY
RFA GOAL #7:
Improve techniques for record linkage and methods for disclosure protection
CODE
R packages
(GitHub, CRAN)
R code, SAS macros
DATA & SUMMARIES
ARTICLES
Articles with Dataverse data citation information
Health outcomes
Simulated health outcomes
Replication instructions
Public & restricted access non-sensitive data
KEY
Secure server
Dataverse
External services

Scientific Questions

1. Is exposure to fine particulate matter ($PM_{2.5}$) below the National Ambient Air Quality Standards (NAAQS) (35 μg/m^3 for daily and 12 μg/m^3 for annual) associated with an increase mortality risks?

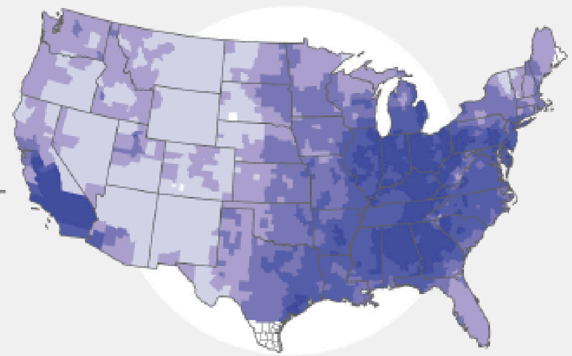
2. Are some populations at higher risk than others?

DATA

- All Medicare participants (n=67,682,479) in the continental United States from 2000 to 2012 (updating the data to 2016)
- Outcomes: all-cause mortality and cause specific hospitalization
- Individual level information: date of death, age of entry, year of entry, sex, race, whether eligible for Medicaid (proxy for socioeconomic status)
- Zip code of residence and other covariates

RESEARCH DATA PLATFORM

EXPOSURES AND INTERVENTIONS (E OR I)



$PM_{2.5}$ exposure levels by county (average 2000-2012)

DATA SOURCES

Criteria air pollutants
EPA AQS daily average of $PM_{2.5}$, ozone, NO_2, 1995-2015;
Daily 1km x 1km predictions of $PM_{2.5}$, ozone, NO_2, 2000-2014

Methane
1km x 1km predictions at 3-day intervals, 2009-present

Weather
NOAA daily estimates (temperature, precipitation, humidity, ...) on a 0.3^0 grid

Power plants
EPA AMPD daily emissions, 1995-2015

Coal mines
MSHA location and producting pits, 1970-2015

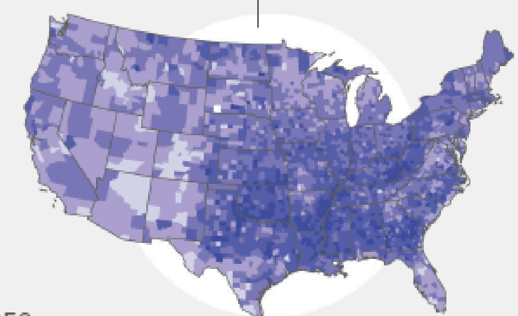
Fracking wells and disposal wells
Drillinginfo database with well location and depth, daily production

Traffic
Annual traffic counts and density from the Department of Transportation

Residential community green space
NASA vegetation index on a 250m^2 grid

Factroies and industrial sites
Geocoded locations of businesses

HEALTH OUTCOMES (Y)



Medicare mortality rate by county (average 2000-2012)

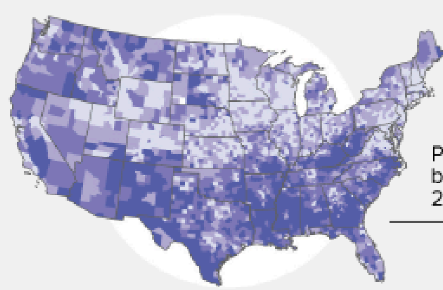
DATA SOURCES

Medicare
28 million per year, 1999-2015

Medicaid
28 million per year, low income, 2010-2011

Aetna
40 million, all ages, above-average income, 2008-2016

CONFOUNDERS (X)



Poverty prevalence by county (average 2000 and 2010)

DATA SOURCES

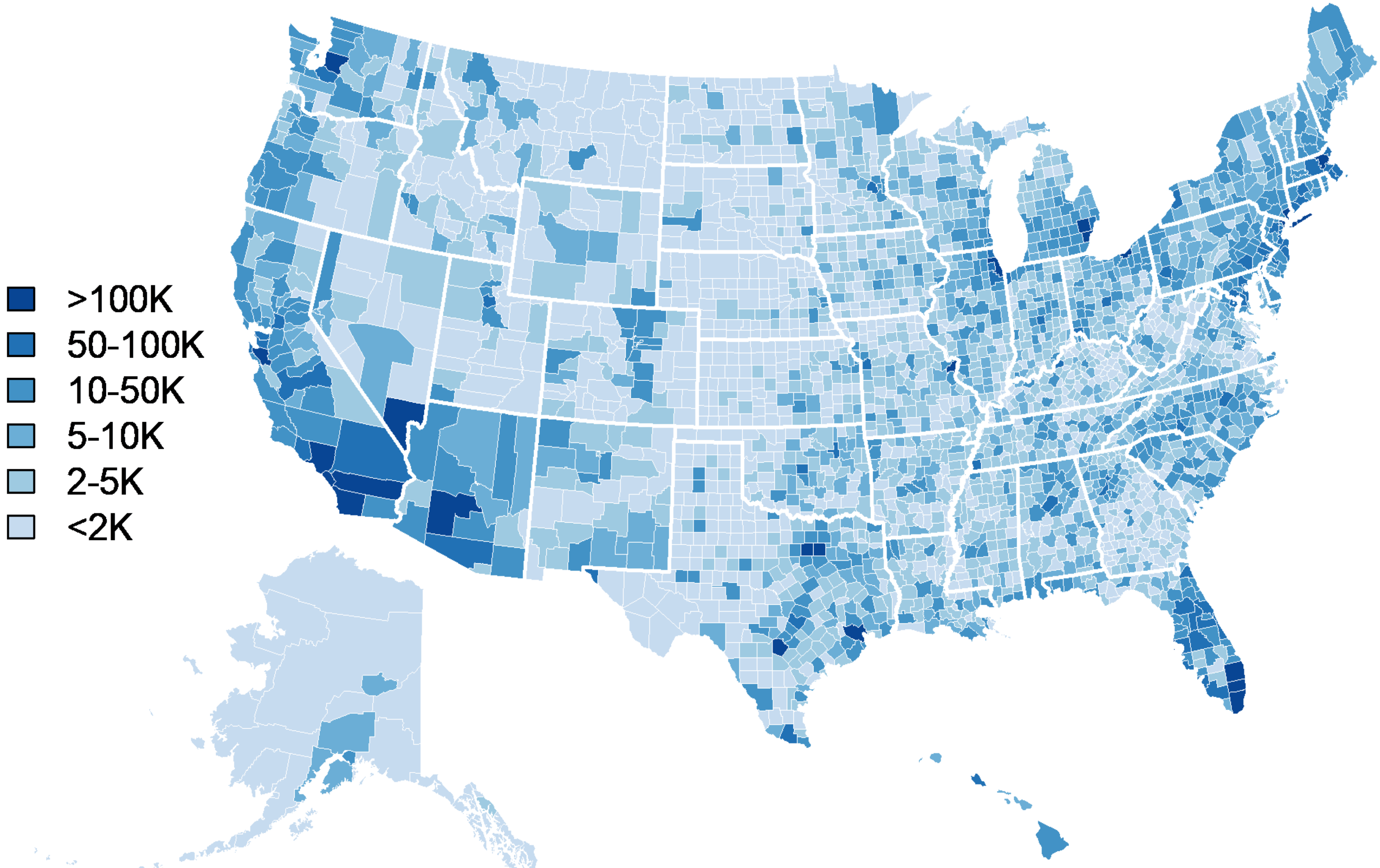
Individual demographics
Age, sex, race, ZIP code of residence

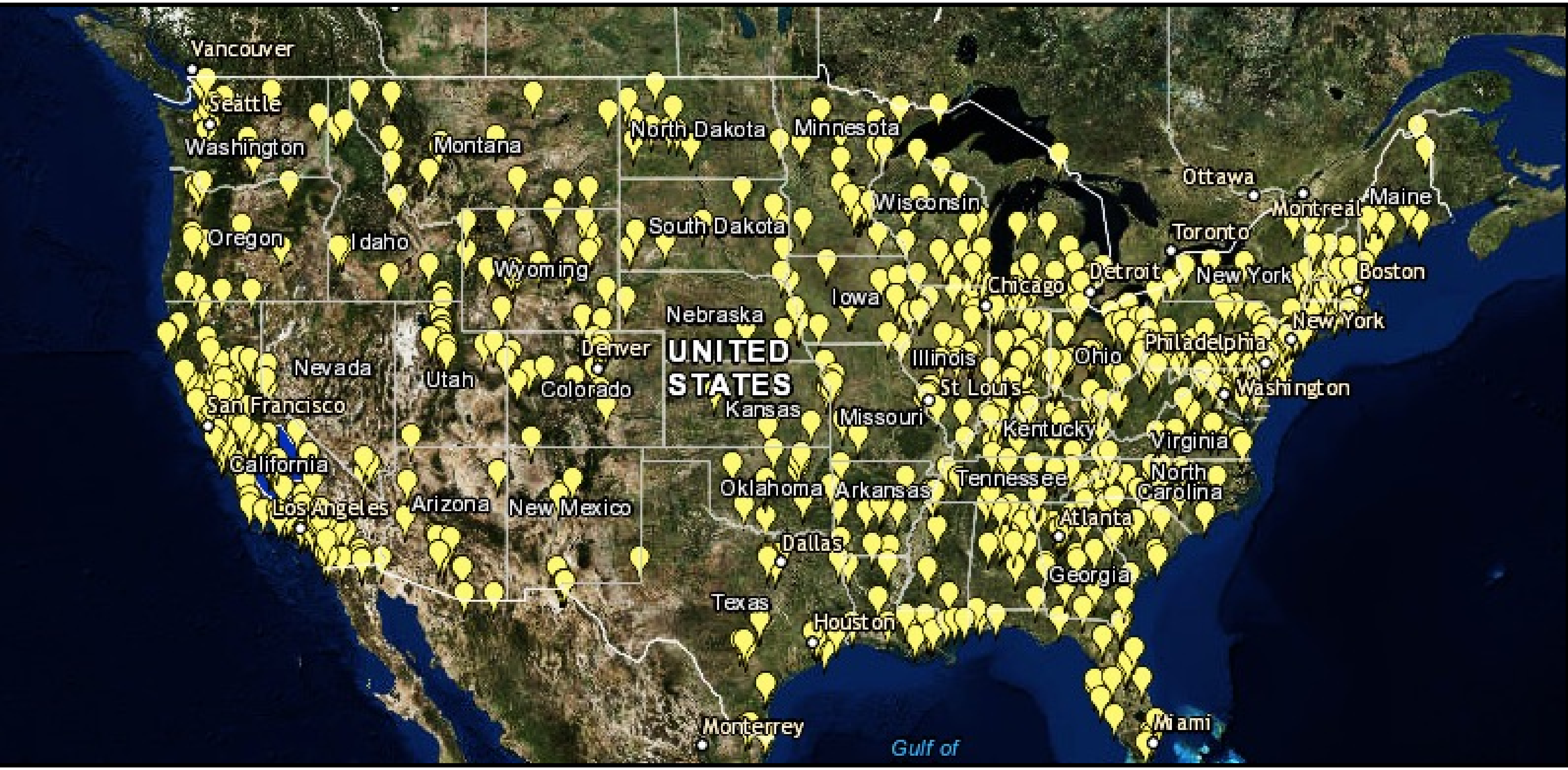
Individual medical history
Previous diagnoses, medications prescribed

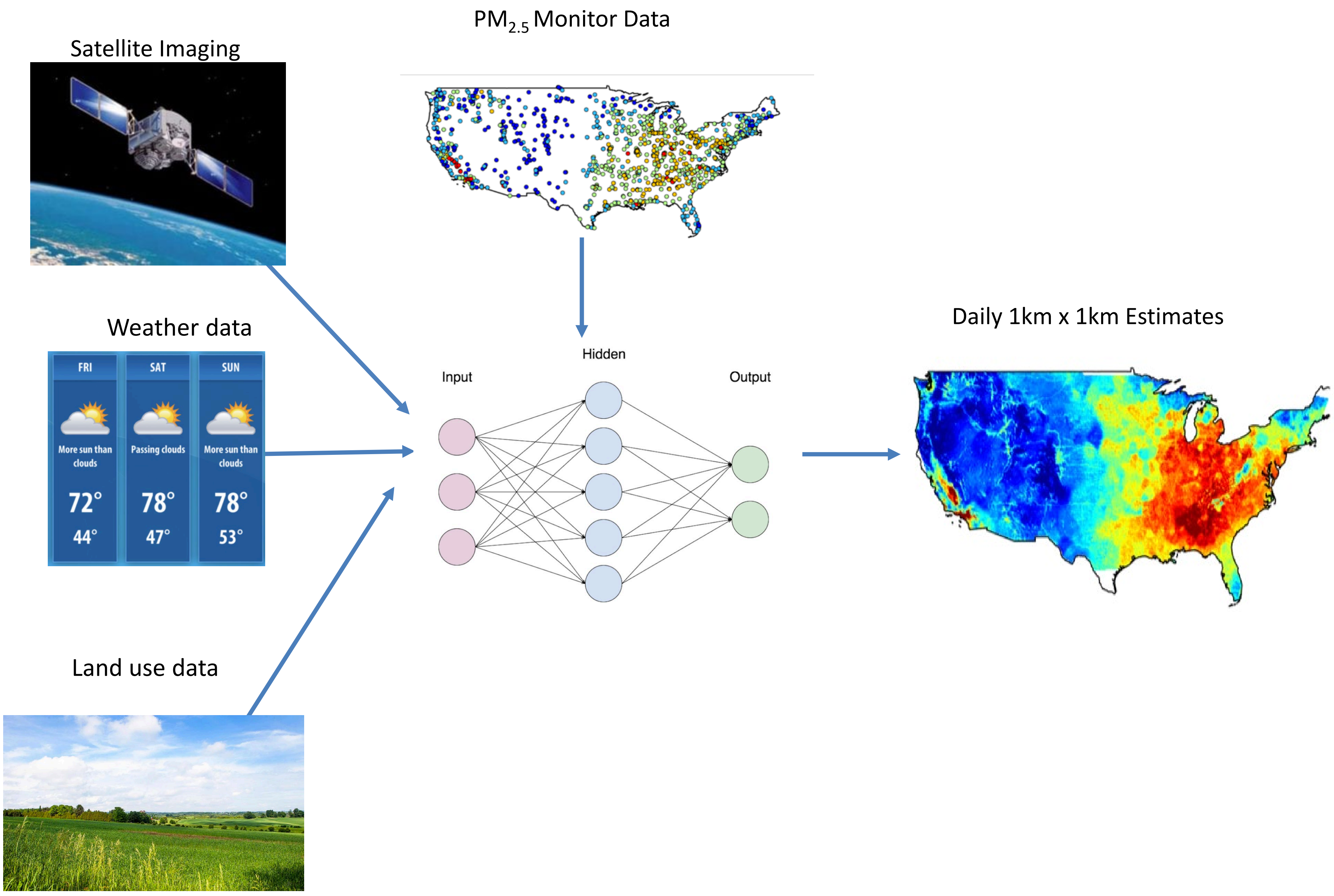
ZIP code level variables
Income, education, demographics, employment, household size

County-level variables
Crime, smoking, BMI

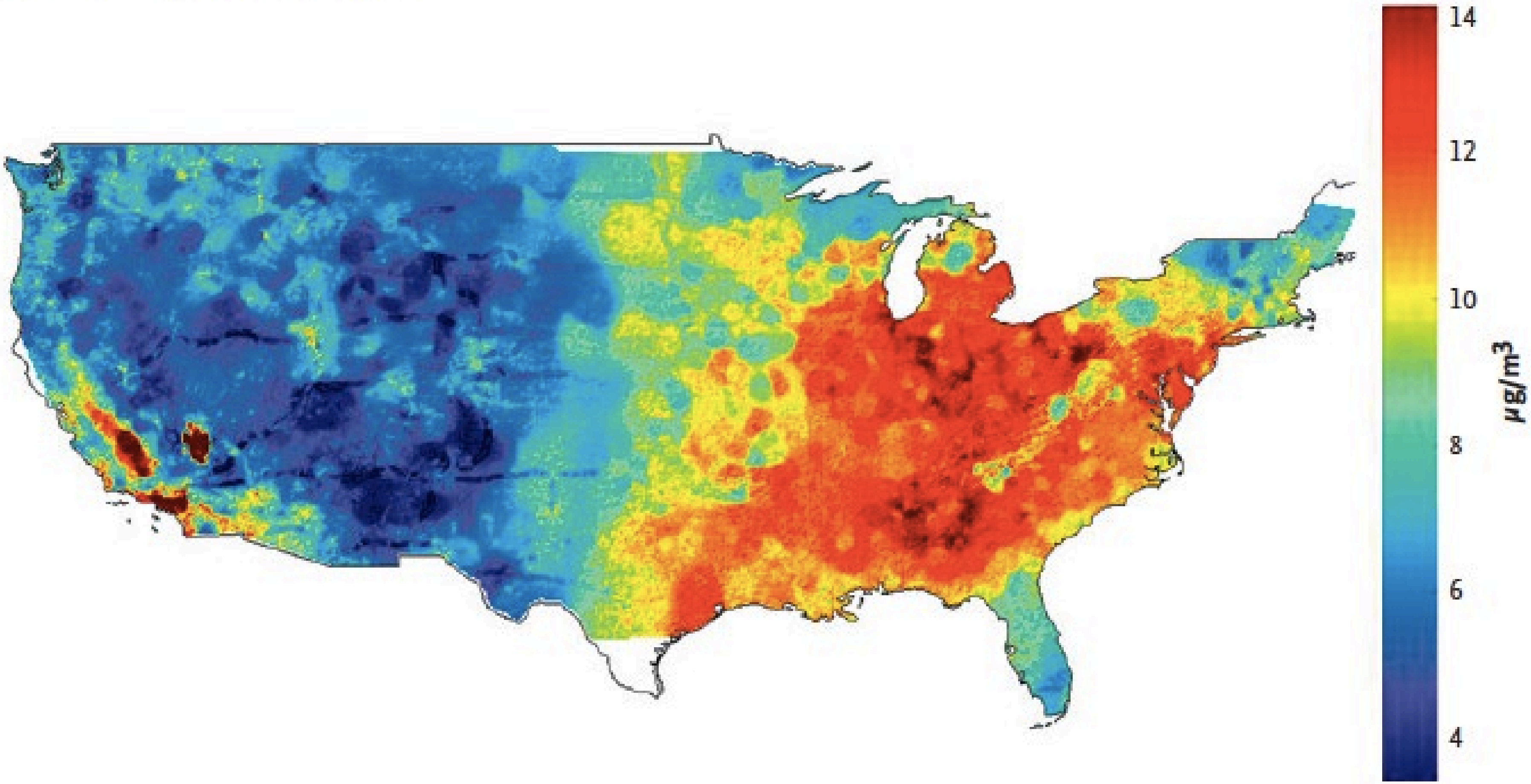
Medicare Data
(open cohort of 60 million enrollees from 1999 to 2012) 460 million person-years




Vancouver
Seattle
Washington
Montana
North Dakota
Minnesota
Ottawa
Maine
Montreal
Wisconsin
South Dakota
Toronto
Oregon
Idaho
Wyoming
Detroit
New York
Boston
Chicago
Iowa
Nebraska
New York
Denver
UNITED
STATES
Illinois
Ohio
Philadelphia
Nevada
Utah
Colorado
St Louis
Washington
San Francisco
Kansas
Missouri
Kentucky
Virginia
California
Oklahoma
Arkansas
Tennessee
North
Carolina
Los Angeles
Arizona
New Mexico
Atlanta
Dallas
Georgia
Texas
Houston
Miami
Monterrey
Gulf of


PM2.5 Monitor Data
Satellite Imaging
Weather data
FRI
SAT
SUN
More sun than clouds
Passing clouds
More sun than clouds
72°
78°
78°
44°
47°
53°
Land use data
Hidden
Input
Output
Daily 1km x 1km Estimates

A Average Concentrations of $PM_{2.5}$



The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JUNE 29, 2017 VOL. 376 NO. 26

Air Pollution and Mortality in the Medicare Population

Qian Di, M.S., Yan Wang, M.S., Antonella Zanobetti, Ph.D., Yun Wang, Ph.D., Petros Koutrakis, Ph.D., Christine Choirat, Ph.D., Francesca Dominici, Ph.D., and Joel D. Schwartz, Ph.D.

Table 1. Cohort Characteristics and Ecologic and Meteorologic Variables.

Characteristic or Variable	Entire Cohort	Ozone Concentration		$PM_{2.5}$ Concentration	
		≥50 ppb*	<50 ppb	≥12 μg/m³	<12 μg/m³
Population					
Persons (no.)	60,925,443	14,405,094	46,520,349	28,145,493	32,779,950
Deaths (no.)	22,567,924	5,097,796	17,470,128	10,659,036	11,908,888
Total person-yr†	460,310,521	106,478,685	353,831,836	212,628,154	247,682,367
Median yr of follow-up	7	7	7	7	7
Average air-pollutant concentrations‡					
Ozone (ppb)	46.3	52.8	44.4	48.0	45.3
$PM_{2.5}$ (μg/m³)	11.0	10.9	11.0	13.3	9.6

Table 2. Risk of Death Associated with an Increase of 10 μg per Cubic Meter in $PM_{2.5}$ or an Increase of 10 ppb in Ozone Concentration.*

Model	$PM_{2.5}$	Ozone
	hazard ratio (95% CI)	
Two-pollutant analysis		
Main analysis	1.073 (1.071–1.075)	1.011 (1.010–1.012)
Low-exposure analysis	1.136 (1.131–1.141)	1.010 (1.009–1.011)
Analysis based on data from nearest monitoring site (nearest-monitor analysis)†	1.061 (1.059–1.063)	1.001 (1.000–1.002)
Single-pollutant analysis‡	1.084 (1.081–1.086)	1.023 (1.022–1.024)

Increases of 10 μg/m^3 in $PM_{2.5}$ and of 10 ppb in ozone were associated with increases in all-cause mortality of 7.3% (95% confidence interval [CI], 7.1 to 7.5) and 1.1% (95% CI, 1.0 to 1.2), respectively.

Adjusted by age, gender, race, previous hospitalization, zip code level income, education etc

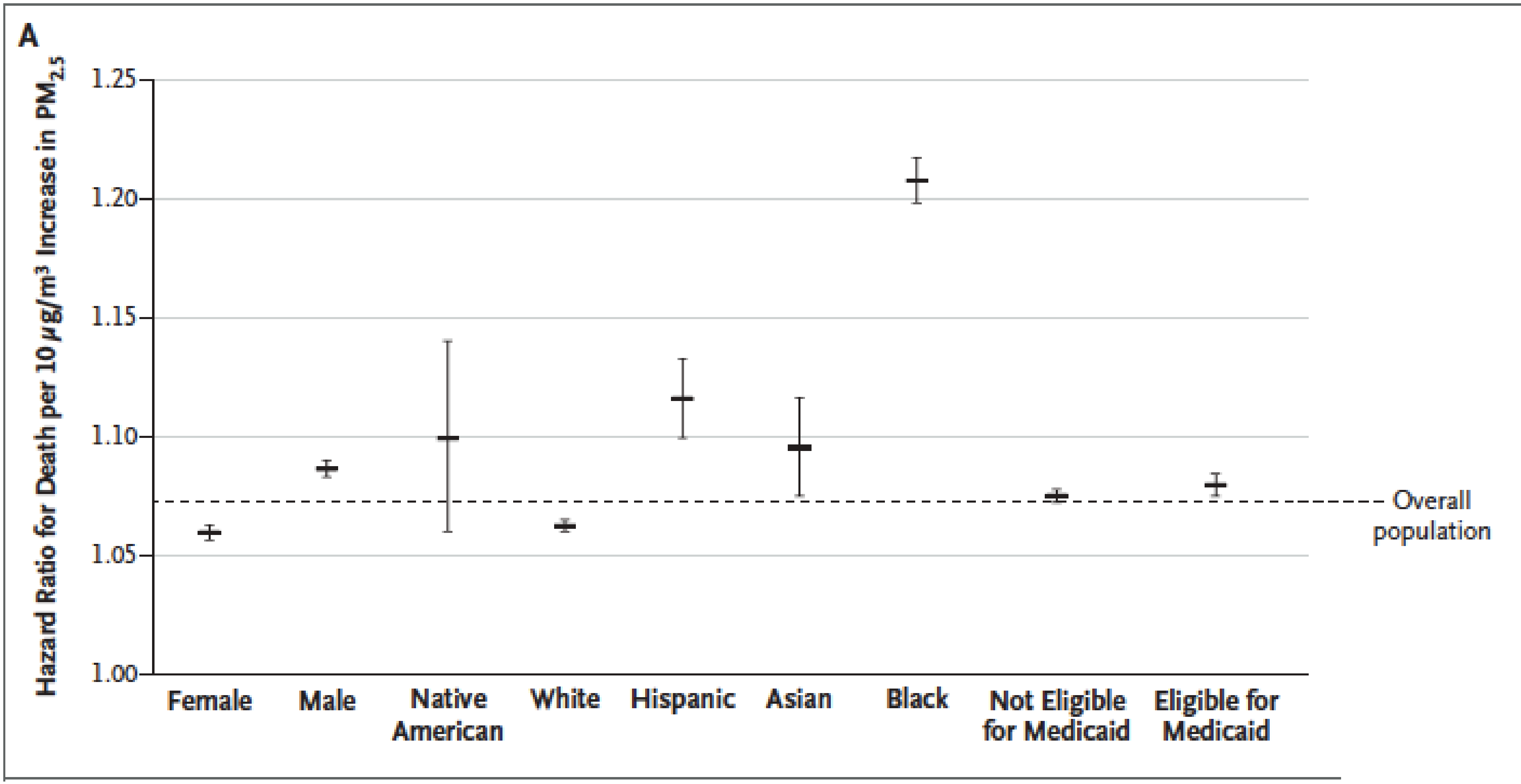


Figure 2. Risk of Death Associated with an Increase of 10 μg per Cubic Meter in $PM_{2.5}$ Concentrations and an Increase of 10 ppb in Ozone Exposure, According to Study Subgroups.

Hazard ratios and 95% confidence intervals are shown for an increase of 10 μg per cubic meter in $PM_{2.5}$ and an increase of 10 parts per billion (ppb) in ozone. Subgroup analyses were conducted by first restricting the population (e.g., considering only male enrollees). The same two-pollutant analysis (the main analysis) was then applied to each subgroup. Numeric results are presented in Tables S3 and S4 in the Supplementary Appendix. Dashed lines indicate the estimated hazard ratio for the overall population.

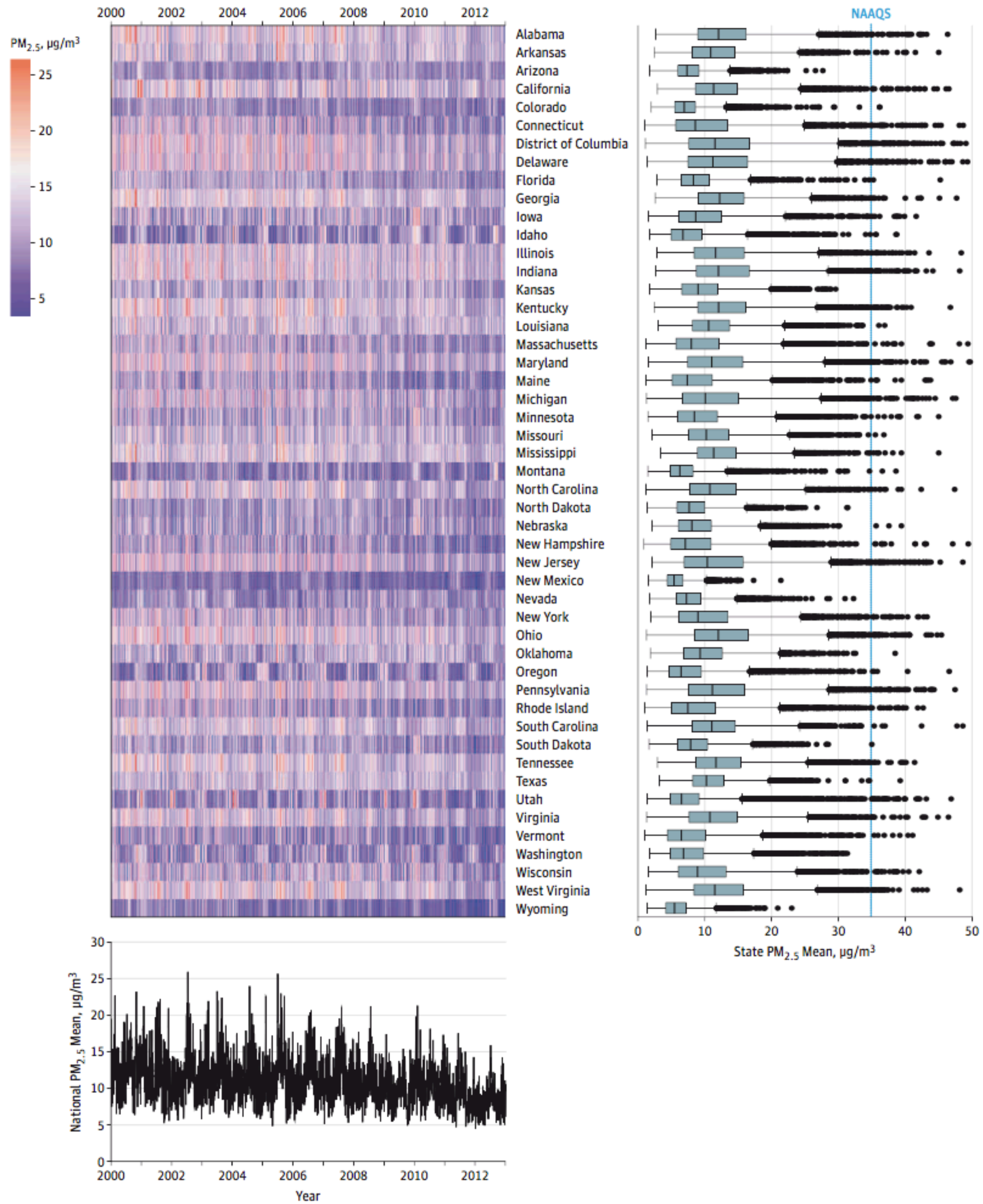
Association of Short-term Exposure to Air Pollution With Mortality in Older Adults

Qian Di, MS; Lingzhen Dai, ScD; Yun Wang, PhD; Antonella Zanobetti, PhD; Christine Choirat, PhD;
Joel D. Schwartz, PhD; Francesca Dominici, PhD

Table 1. Baseline Characteristics of Study Population (2000-2012)

Baseline Characteristic	Value
Case days, No.	22 433 862
Control days, No.	76 143 209
Among All Cases (n = 22 433 862), %	
Age at death, y	
≤69	10.38
70-74	13.37
75-84	38.48
≥85	37.78
Sex	
Male	44.73
Female	55.27
Race/ethnicity	
White	87.34
Black	8.87
Asian	1.03
Hispanic	1.51
Native American	0.31
Medicaid Eligibility (n = 22 433 862), %	
Ineligible	77.36
Eligible	22.64

Figure 1. Daily Mean $PM_{2.5}$ Concentrations in the Continental United States, 2000-2012



Daily mean fine particulate matter ($PM_{2.5}$) concentrations were calculated and plotted by state. The time-series plot at the bottom indicates the national daily mean values across all locations. Boxplots show the distribution of daily $PM_{2.5}$ levels for each state. The blue dashed line indicates the daily National Ambient Air Quality Standards (NAAQS) for $PM_{2.5}$ (35 μg/m^3). The line across the box, upper hinge, and lower hinge represent the median value, 75th percentile (Q3), and 25th percentile (Q1), respectively. The upper whisker is located at the smaller of the maximal value and Q3 + 1.5 × interquartile range; the lower whisker is located at the larger of the minimal value and Q1 - 1.5 × interquartile range. Any values that lie beyond the upper and lower whiskers are outliers.

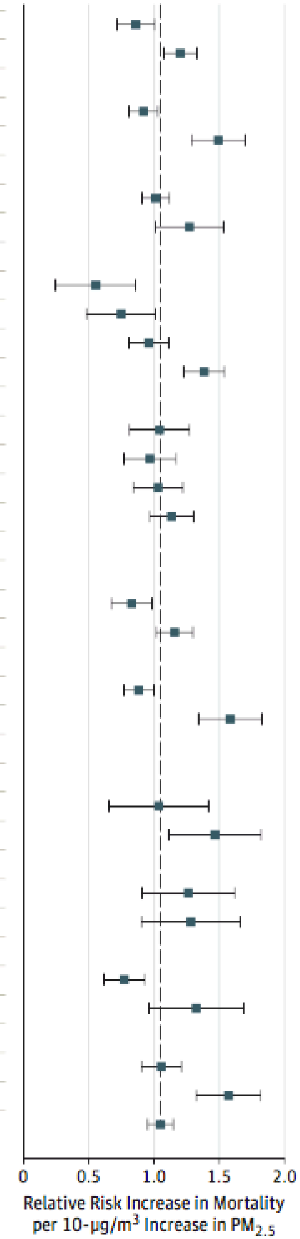
Table 2. Relative Risk Increase and Absolute Risk Difference of Daily Mortality Associated With Each 10-μg/m^3 Increase in $PM_{2.5}$ and Each 10-ppb Increase in Ozone

	Relative Risk Increase, % (95% CI)		Absolute Risk Difference in Daily Mortality Rates, No. per 1 Million Persons at Risk per Day (95% CI)[a]	
Air Pollutant Analysis	$PM_{2.5}$	Ozone[b]	$PM_{2.5}$	Ozone[b]
Main analysis[c]	1.05 (0.95-1.15)	0.51 (0.41-0.61)	1.42 (1.29-1.56)	0.66 (0.53-0.78)
Low-exposure analysis[d]	1.61 (1.48-1.74)	0.58 (0.46-0.70)	2.17 (2.00-2.34)	0.74 (0.59-0.90)
Single-pollutant analysis[e]	1.18 (1.09-1.28)	0.55 (0.48-0.62)	1.61 (1.48-1.73)	0.71 (0.62-0.79)
Nearest monitors analysis[f]	0.83 (0.73-0.93)	0.35 (0.28-0.41)	1.13 (0.99-1.26)	0.45 (0.37-0.53)

Time stratified case crossover design (Lumley, Sheppard):

- Case day was defined as the date of death, for the same person, we compared daily air pollution exposure on the case day vs daily air pollution exposure on control days
- Control days were chosen (1) on the same day of the week as the case day to control for potential confounding effect by day of week; (2) before and after the case day (bidirectional sampling) to control for time trend; and (3) only in the same month as the case day to control for seasonal and sub-seasonal patterns
- Individual-level covariates and zip code–level covariates that did not vary day to were not considered to be confounders as they remain constant when comparing case days vs control days.

Model	Relative Risk Increase in Mortality per 10-μg/m^3 Increase in $PM_{2.5}$, % (95% CI)	*P* Value for Effect Modification
Sex		
Male	0.86 (0.72-1.00)	[Reference]
Female	1.20 (1.07-1.33)	<.001[a]
Medicaid eligibility		
Noneligible	0.92 (0.81-1.03)	[Reference]
Eligible	1.49 (1.29-1.70)	<.001[a]
Race/ethnicity		
White	1.01 (0.91-1.12)	[Reference]
Nonwhite	1.27 (1.01-1.53)	.07
Age, y		
≤69	0.55 (0.25-0.86)	[Reference]
70-74	0.75 (0.48-1.01)	.35
75-84	0.96 (0.80-1.11)	.02[a]
≥85	1.38 (1.23-1.54)	<.001[a]
Population density		
Low	1.04 (0.81-1.27)	[Reference]
Medium low	0.97 (0.76-1.17)	.64
Medium high	1.03 (0.84-1.22)	.95
High	1.13 (0.97-1.30)	.52
Whites		
Sex		
Male	0.83 (0.67-0.99)	[Reference]
Female	1.16 (1.02-1.30)	.002[a]
Medicaid eligibility		
Noneligible	0.88 (0.77-1.00)	[Reference]
Eligible	1.58 (1.34-1.83)	<.001[a]
Nonwhites		
Sex		
Male	1.03 (0.65-1.42)	[Reference]
Female	1.47 (1.12-1.82)	.01
Medicaid eligibility		
Noneligible	1.26 (0.91-1.62)	[Reference]
Eligible	1.28 (0.90-1.66)	.94
Medicaid eligibility, males		
Noneligible	0.77 (0.61-0.93)	[Reference]
Eligible	1.32 (0.96-1.69)	.006
Medicaid eligibility, females		
Noneligible	1.06 (0.90-1.21)	[Reference]
Eligible	1.57 (1.32-1.82)	<.001[a]
Overall	1.05 (0.95-1.15)	



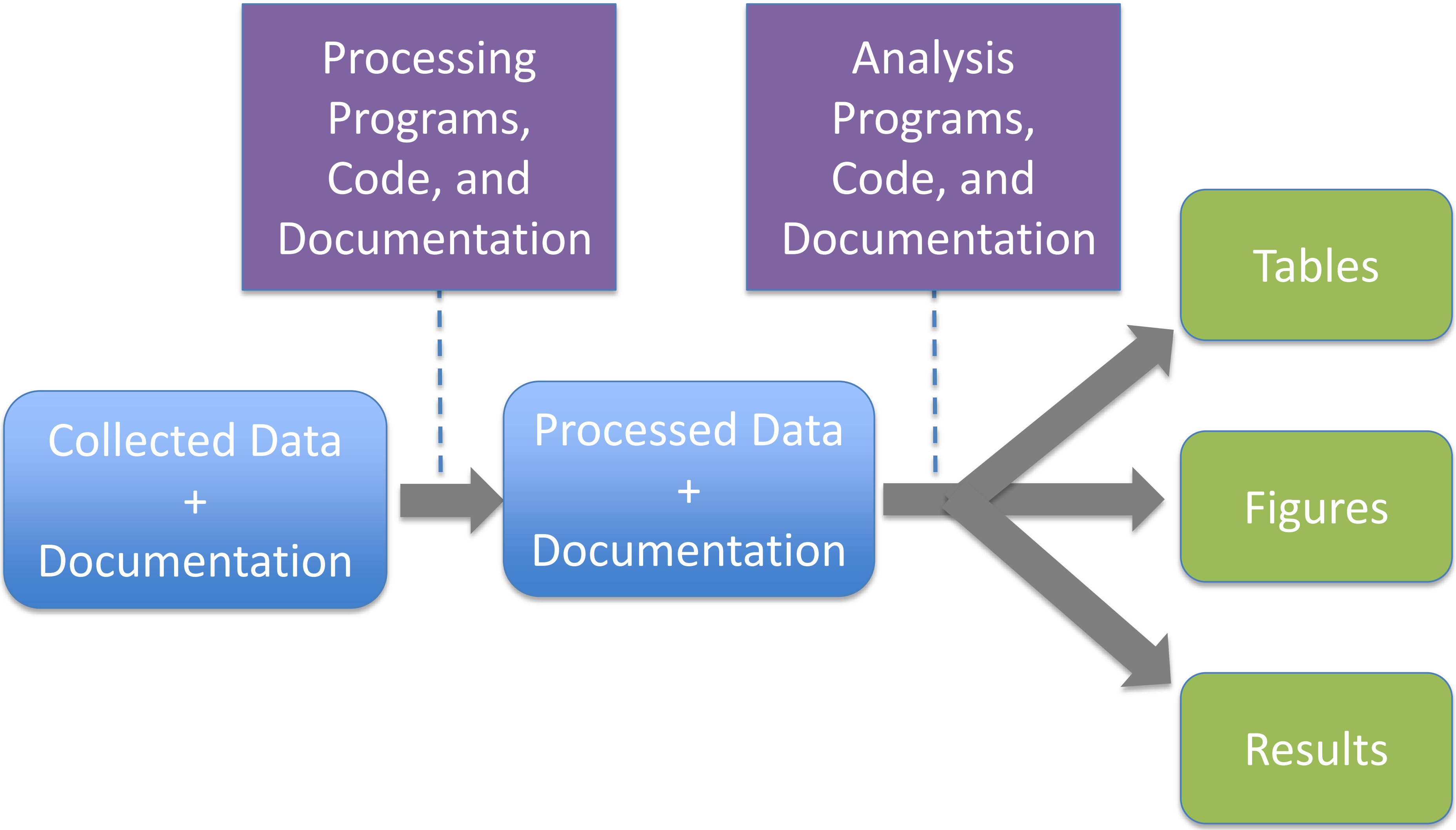
Methodological challenges

- **Methods**
- Wu X et al (2019) . Matching on Generalized Propensity Scores with Continuous Exposures. https://arxiv.org/pdf/1812.06575.pdf (submitted)
- **Exposure error correction**
- Wu X et al . (2018) Causal inference in the context of an error prone exposure: air pollution and mortality. The *Annals of Applied Statistics*
- **Unmeasured confounding bias**
- Makar M et al. (2017). Estimating the Causal Effect of Lowering Particulate Matter Levels below the United States Standards on Hospitalization and Death. *Epidemiology.*
- **Discovery of heterogeneous subgroups to characterize vulnerability**
- Lee K et al 2019 Discovering Effect Modification and Randomization in Air Pollution Studies. *Journal of American Statistical Association* https://arxiv.org/pdf/1802.06710.pdf
- **Reproducibility**
- airpred: A Flexible R Package Implementing Methods for Predicting Air Pollution (https://arxiv.org/abs/1805.11534)

Planned and Ongoing Analyses

- Exposure estimation for PM, Ozone, and NO_2 to 2016 via ensemble learning (poster)
- Applying causal inference methods to same data used to complete prospective national Medicare cohort study
- Discovering heterogenous groups under a causal inference framework
- Case crossover study of Medicaid data
- Reproducibility of the Research Data Platform

Reproducible workflows



The Team


Marianthi Kioumourtzoglou


Joel Schwartz


Joey Antonelli


Antonella Zanobetti


Iste Zahn


Corwin Zigler


Christine Choirat

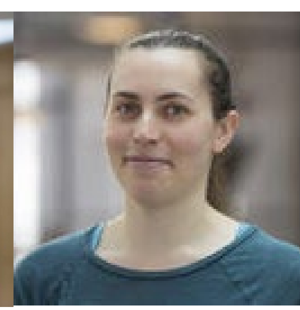

Rachel Nethery


Qian Di


Kwonsang Lee


Qiao Wu


Georgia Papadogeorgou


Danielle Braun


Ben Sabath


Emma Thomas