SUPPLEMENTAL MATERIAL

Data S1.

Expanded Methods

Geocoding Procedure:

We obtained patient addresses from two medical records datasets: the Epic® electronic health record (EHR) dataset contained 112,519 addresses from visits occurring from 2014 to 2016; the homegrown WebCIS (legacy) EHR dataset contained 63,856 addresses from visits prior to 2014. We began cleaning data by standardizing variable formats. The variables included two text variables for address lines 1 (street address) and 2 (apartment or unit number), a text variable for city, a text variable for state, and a numeric variable for zip/postal code. We then removed duplicate addresses: the exact same text entered as an address two or more times for a single participant ("Main St" and "Main Street" would not be considered duplicates at this stage). This reduced our number of addresses to 86,616 in the Epic® EHR dataset and 58,905 in legacy. Both datasets underwent initial geocoding using proc geocode in SAS version 9.4 (SAS Institute Inc. Cary, NC). We matched addresses to latitude and longitude using street lookup data from SAS MapsOnline, which was generated from the US Census Bureau's TIGER/Line shapefiles. SAS proc geocode detects and corrects for minor textual variants, such as common spelling errors and abbreviations. Epic® and legacy data were then merged based on level of initial geocoding match: street level (n = 101 594), zip code level (n = 43 285), city level (n = 500), or unmatched (n = 142).

We then cleaned data to improve the level of matching for addresses not matched to the street level, where possible. We began by cleaning data with addresses that were unmatched. Various

indicators of null address elements (“unknown”, “none”, “NA”, etc.) were standardized to all read “NULL”. Then, observations were separated into those with no valid address information and those with partial information (i.e. only zip code). Those with partial address information were merged to the city matched dataset, while those with no valid address information were kept as unmatched.

For those addresses in the city matched dataset, we standardized null address elements as above, standardized the spelling of PO Box, and removed punctuation from text fields. PO Boxes are Post Office Boxes, for individuals who receive mail at a Post Office instead of their home. PO Box addresses can be geocoded to zip code, but not street address. We assumed that PO Boxes were located in the same zip code where a patient resided. We then identified likely mistakes in which the listed city was a near but not exact match to the listed zip code and corrected the zip code. We then re-geocoded city matched addresses. Those who matched to zip code at this point were merged with the zip code matched dataset.

For those addresses matched to zip code, we standardized null address elements as above, standardized the spelling of PO Box, removed punctuation, corrected obvious misspellings and unconventional abbreviations, and standardized indicators of homelessness. Many individuals experiencing homelessness have listed a zip code where they typically stay, although they may be transient. Some participants listed both a PO Box and a street address. For those individuals, we attempted to geocode the street address. We then standardized text fields of data entries including apartments. If street address and apartment number were entered in an unconventional format (i.e. listing apartment number before street address), we attempted to extract the street

address and re-geocode the address. Then, for any addresses with text on the second address line, we attempted to geocode the second address line. Any addresses that were re-geocoded to street level at this stage were merged to street level matches.

For street level matches, we corrected the address of a common senior living community that geocoded incorrectly in some instances. At this point, we removed duplicate addresses: multiple addresses matched to the same street (for street level matches) or zip code (for zip code level matches) for the same participant. We then matched participants to their "best" address(es), by source and by matching level. We created a dummy dataset with one observation per participant (n = 41 913). We first matched participants to addresses from Epic® that were street matches. If there were multiple unique addresses per participant, we kept all addresses (n = 32 614 addresses). If a participant was not matched to street in Epic™, we repeated the same procedure for addresses from Epic® that were zip code matches (n = 3 857 addresses). For participants still not matched to an address, we matched them to addresses from legacy files that were street matched (n = 9 006 addresses), keeping only the most recent based on account number, an approximate estimation of chronology. For participants still not matched, we repeated the procedure in order for the following sources and types of match, keeping only the most recent instance: legacy zip code match (n = 2 118 addresses), Epic® city match (n=47 addresses), legacy city match (n = 2 addresses), and unmatched (n = 17 participants).

Once participants' address(es) were established, we endeavored to match participants to their nearest EPA $PM_{2.5}$ monitor. We limited to participants who resided in NC and matched addresses to all visit dates. For each visit date, we calculated the date 365 prior to the visit. We then took

all data from EPA monitors across the state of NC during the period from January 1, 2003 until December 31, 2016. We calculated the first and last dates within that period when each monitor had valid data. We matched all monitors that were active during the entire 365 days prior to participants' visits to their addresses. We calculated the distances between the monitor and the addresses using the latitude and longitude of the addresses and monitors and the geodist function in SAS 9.4. We then kept the nearest monitor to each address for each visit date. We then averaged the daily mean $PM_{2.5}$ value for the 365 days prior to each visit.

Co-morbid disease ICD-9 and ICD-10 codes:

International Classification of Disease (ICD)-9 and ICD-10 codes used to determine existing co-morbidities are given below. A * is used as a wildcard character which was used to capture all subcodes for relevant disease definitions.

Type 2 diabetes: 250, 250.0, 250.00, 250.02, 250.1, 250.10, 250.12, 250.2, 250.20, 250.22, 250.3, 250.30, 250.32, 250.4, 250.40, 250.42, 250.5, 250.50, 250.52, 250.6, 250.60, 250.62, 250.7, 250.70, 250.72, 250.8, 250.80, 250.82, 250.9, 250.90, 250.92, E11.*

Hypertension: 401*, I10*

Dyslipidemia: 272*, E78*

Peripheral Arterial Disease: 443*, I73*

Chronic Obstructive Pulmonary Disorder (bronchitis): 491*, J44*

Emphysema: 492*, J43*

Ischemic Heart Disease: 414*, I20, I21, I22, I23, I24, I25

Table S1. Associations between all-cause mortality and $PM_{2.5}$.

Stratification	Model	HR	LCI	UCI	N	N deaths
None	basic	1.12	1.10	1.15	23302	4496
None	full	1.13	1.10	1.15	23012	4445
Males	basic	1.12	1.09	1.16	11224	2229
Males	full	1.13	1.09	1.16	11075	2208
Females	basic	1.12	1.09	1.16	12078	2267
Females	full	1.12	1.09	1.16	11937	2237
European Americans	basic	1.11	1.09	1.14	14216	2953
European Americans	full	1.12	1.09	1.15	14090	2926
African Americans	basic	1.13	1.09	1.18	7063	1322
African Americans	full	1.13	1.08	1.18	6934	1300
<50 y	basic	1.10	1.01	1.19	3253	383
<50 y	full	1.09	1.00	1.18	3182	376
50-65 y	basic	1.11	1.06	1.16	6943	1126
50-65 y	full	1.11	1.06	1.16	6825	1110
≥65 y	basic	1.13	1.11	1.16	13106	2987
≥65 y	full	1.14	1.11	1.16	13005	2959
$PM_{2.5} < 12$ μg/m^3	basic	1.11	1.07	1.14	18268	2166
$PM_{2.5} < 12$ μg/m^3	full	1.11	1.08	1.15	18055	2151
Type 2 diabetes	basic	1.12	1.08	1.16	7853	1768
Type 2 diabetes	full	1.12	1.08	1.16	7741	1743
PAD	basic	1.12	1.08	1.15	8195	1767
PAD	full	1.12	1.08	1.16	8100	1749
COPD	basic	1.13	1.09	1.17	8293	1772
COPD	full	1.13	1.10	1.17	8177	1747
Dyslipidemia	basic	1.13	1.10	1.16	17215	3402
Dyslipidemia	full	1.13	1.10	1.16	17017	3367
Hypertension	basic	1.12	1.09	1.15	17027	3482
Hypertension	full	1.12	1.10	1.15	16815	3442
IHD	basic	1.13	1.10	1.16	13260	2871
IHD	full	1.13	1.10	1.16	13093	2842

Associations are given per 1 μg/m^3 increase in $PM_{2.5}$. Associations were stratified on sex, race, Hispanic ethnicity, age at heart failure diagnosis, exposure level, and pre-exiting co-morbidities. The "basic" model adjusted for age, race, sex, and distance to the nearest monitor while the "full"

model adjusted for age, sex, race, distance to the nearest monitor, median income, median house value, percent of individuals on public assistance, urbanicity, and percent of households below federal poverty line.

* COPD = chronic obstructive pulmonary disorder; HR = hazard ratio; IHD = Ischemic heart disease; LCI = lower 95% confidence interval; PAD = peripheral arterial disease; UCI = upper 95% confidence interval

Table S2. Associations between all-cause mortality and $PM_{2.5}$ assessed using modeled $PM_{2.5}$ data at 1x1 km resolution.

Stratification	Model	HR	LCI	UCI	N	N deaths
None	basic	1.18	1.15	1.21	23127	4480
None	full	1.19	1.16	1.22	22841	4429
Males	basic	1.17	1.13	1.21	11146	2219
Males	full	1.18	1.14	1.22	11001	2198
Females	basic	1.19	1.15	1.23	11981	2261
Females	full	1.20	1.15	1.24	11840	2231
European Americans	basic	1.17	1.14	1.21	14113	2945
European Americans	full	1.18	1.15	1.22	13988	2918
African Americans	basic	1.18	1.13	1.24	7004	1315
African Americans	full	1.19	1.13	1.24	6877	1293
<50 y	basic	1.17	1.06	1.28	3214	381
<50 y	full	1.16	1.05	1.27	3146	374
50-65 y	basic	1.17	1.11	1.23	6883	1124
50-65 y	full	1.18	1.12	1.24	6766	1108
≥65 y	basic	1.19	1.15	1.22	13030	2975
≥65 y	full	1.19	1.16	1.23	12929	2947
$PM_{2.5}$ < 12 µg/m^3	basic	1.18	1.14	1.22	18109	2152
$PM_{2.5}$ < 12 µg/m^3	full	1.19	1.15	1.24	17900	2137
T2D	basic	1.18	1.14	1.23	7812	1765
T2D	full	1.19	1.14	1.24	7702	1740
PAD	basic	1.16	1.12	1.21	8135	1760
PAD	full	1.17	1.13	1.22	8042	1742
COPD	basic	1.19	1.15	1.24	8241	1767
COPD	full	1.20	1.15	1.25	8127	1742
Dyslipidemia	basic	1.18	1.15	1.21	17093	3391
Dyslipidemia	full	1.19	1.15	1.22	16899	3356
Hypertension	basic	1.17	1.13	1.20	16903	3468
Hypertension	full	1.18	1.14	1.21	16694	3428
IHD	basic	1.18	1.14	1.21	13169	2863
IHD	full	1.19	1.15	1.22	13003	2834
rural	full	1.19	1.14	1.24	7634	1534
urban	full	1.22	1.17	1.27	8840	1697

Associations are given per 1 $\mu g/m^3$ increase in $PM_{2.5}$ with PM2.5 as assessed using the ensemble model values which have a 1x1 km resolution. Associations were stratified on sex, race, Hispanic ethnicity, age at heart failure diagnosis, exposure level, pre-exiting co-morbidities, and urban (100% urban block group) vs rural (bottom 3^{rd} of urbanicity distribution; ~ < 39% urban) residence based on the 2000 Census block group urbanicity assessment. The "basic" model adjusted for age, race, sex, and distance to the nearest monitor while the "full" model adjusted for age, sex, race, distance to the nearest monitor, median income, median house value, percent of individuals on public assistance, urbanicity, and percent of households below federal poverty line.

* COPD = chronic obstructive pulmonary disorder; HR = hazard ratio; IHD = Ischemic heart disease; LCI = lower 95% confidence interval; PAD = peripheral arterial disease; UCI = upper 95% confidence interval

Table S3. Sensitivity analyses based on visit and age at end of observation.

Primary Analysis; HR (CI)	pre 1/1/2015; HR (CI)	Age Outlier Removed; HR (CI)	Age at end < 100; HR (CI)	At least 1 prior visit; HR (CI)
1.13 (1.10, 1.15)	1.10 (1.07, 1.12)	1.13 (1.10, 1.15)	1.13 (1.11, 1.16)	1.11 (1.08, 1.14)

Due to possible bias from lack of observation of deaths, which would result in individuals being "followed" indefinitely, we performed sensitivity analyses based on the age at the end of observation by removing outliers ($\pm$3*standard deviation; Age Outlier removed) and restricting to those < 100 years old at the end of observation time (Age at end < 100). We also performed a sensitivity analyses restricting to individuals seen before 1/1/2015 (pre 1/1/2015) as more individuals were seen in later years but due to the short follow up time would have had a much lower probability of death. Finally, we examined a sensitivity analysis restricting to individuals with at least one prior visit (At least 1 prior visit). No differences were seen across any of the sensitivity analyses. The full adjustment model was used. Also shown is are the results from the primary analysis for comparison.

* CI = 95% confidence interval; HR = hazard ratio

Table S4. Clinical covariates for individuals with diastolic and systolic heart failure.

	Diastolic HF (N = 6385)			Systolic HF (N = 7120)		
	Mean	SD	IQR	Mean	SD	IQR
Age (y)	69.7	14.0	20.7	64.6	15.1	21.7
Follow-up time (y)	4.3	3.12	4.07	3.69	2.69	3.06
Distance to monitor (km)	20.0	14.3	19.9	22.0	15.1	21.1
Monitor PM2.5 (ug/m3)	10.2	2.14	3.54	9.69	1.90	3.00
Model PM2.5 (ug/m3)	10.4	1.72	2.62	9.95	1.45	1.40
Households below federal poverty line (%)	18.0	14.9	18.6	18.5	14.8	19.5
Median Home Value ($)	190260	110753	122100	173793	105746	111900
Median Household Income ($)	54092	26870	30996	51707	26053	29709
Urbanicity (%)	64.1	41.7	83.6	61.1	42.5	94.8
Households receiving public assistance (%)	1.98	3.06	2.91	2.00	2.97	2.91
	N	%		N	%	
African American	1987	31.1		2177	30.6	
European American	3982	62.4		4341	61.0	
Other Race	416	6.5		602	8.5	
Male	2285	35.8		4379	61.5	
Within 30 km of monitor	4992	78.2		5255	73.8	
Death (All Cause)	1465	22.9		1062	14.9	
Type 2 diabetes	2574	40.3		2004	28.1	
IHD	3833	60.0		4214	59.2	
COPD	2786	43.6		2159	30.3	
PAD	2677	41.9		2226	31.3	
Hypertension	5315	83.2		4714	66.2	
Dyslipidemia	5303	83.1		5177	72.7	

BMI = body mass index; COPD = chronic obstructive pulmonary disorder; CVD = cardiovascular disease; IHD = Ischemic heart disease; PAD = peripheral arterial disease; SD = standard deviation

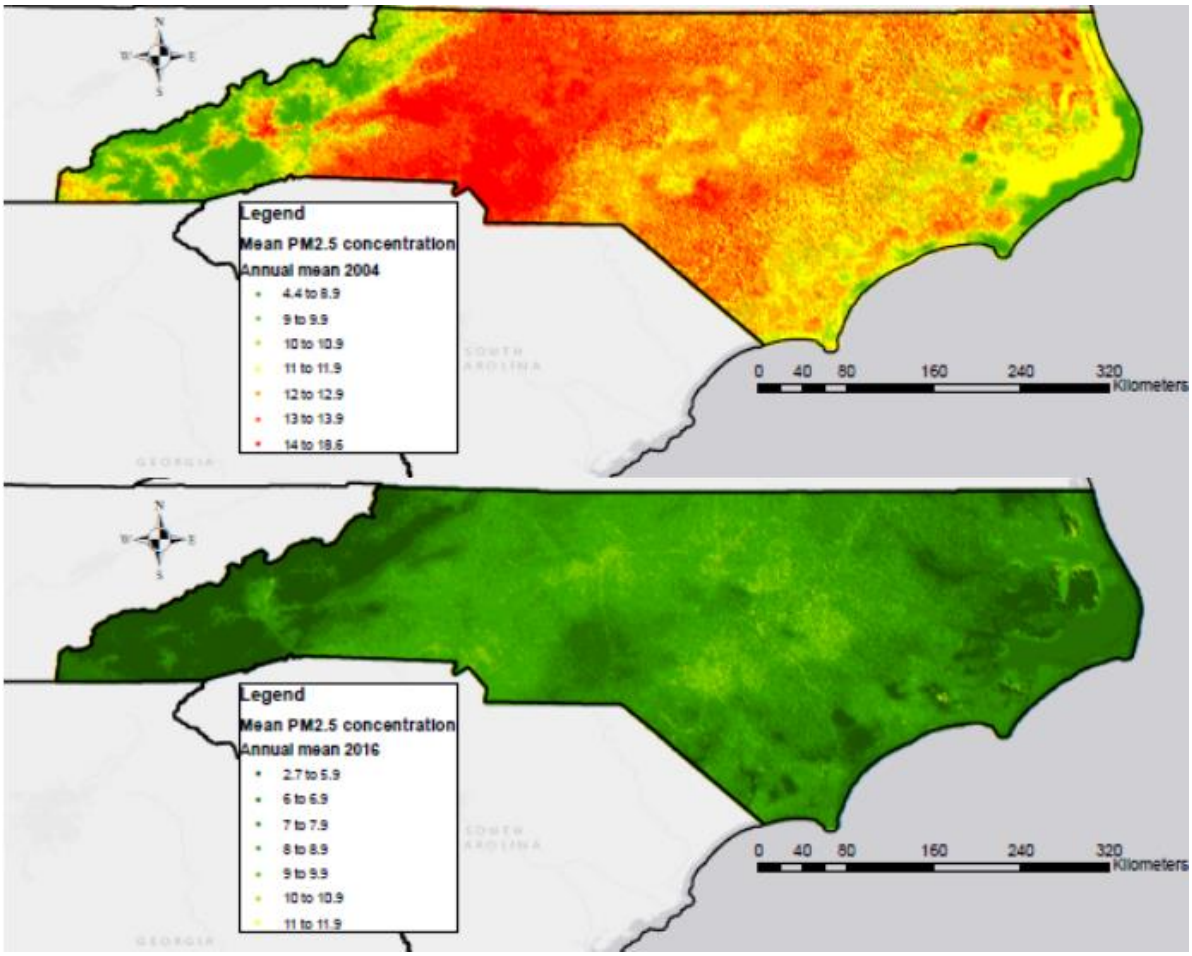
Table S5. Associations between all-cause mortality and binary $PM_{2.5}$ cutoff.

Stratification	Model	HR	LCI	UCI	N	N deaths
None	basic	1.79	1.68	1.91	23302	4496
None	full	1.77	1.66	1.89	23012	4445
Males	basic	1.76	1.61	1.93	11224	2229
Males	full	1.75	1.60	1.92	11075	2208
Females	basic	1.83	1.67	2.00	12078	2267
Females	full	1.79	1.64	1.96	11937	2237
European Americans	basic	1.78	1.64	1.92	14216	2953
European Americans	full	1.76	1.63	1.91	14090	2926
African Americans	basic	1.81	1.61	2.03	7063	1322
African Americans	full	1.76	1.57	1.98	6934	1300
<50 y	basic	2.03	1.63	2.51	3253	383
<50 y	full	1.96	1.58	2.44	3182	376
50-65 y	basic	1.74	1.53	1.97	6943	1126
50-65 y	full	1.72	1.51	1.96	6825	1110
≥65 y	basic	1.78	1.65	1.93	13106	2987
≥65 y	full	1.76	1.63	1.91	13005	2959
Type 2 diabetes	basic	1.80	1.63	1.99	7853	1768
Type 2 diabetes	full	1.77	1.60	1.96	7741	1743
PAD	basic	1.63	1.47	1.80	8195	1767
PAD	full	1.63	1.47	1.80	8100	1749
COPD	basic	1.83	1.66	2.02	8293	1772
COPD	full	1.81	1.64	2.00	8177	1747
Dyslipidemia	basic	1.78	1.65	1.91	17215	3402
Dyslipidemia	full	1.75	1.63	1.88	17017	3367
Hypertension	basic	1.78	1.65	1.91	17027	3482
Hypertension	full	1.75	1.63	1.88	16815	3442
IHD	basic	1.79	1.66	1.94	13260	2871
IHD	full	1.77	1.64	1.92	13093	2842

Associations are given based on a binary indicator for annual average $PM_{2.5}$ exposure < 12 $\mu g/m^3$ (the current EPA annual average standard for $PM_{2.5}$). Associations were stratified on sex, race, Hispanic ethnicity, age at HF diagnosis, exposure level, and pre-exiting co-morbidities.

* COPD = chronic obstructive pulmonary disorder; HR = hazard ratio; IHD = Ischemic heart disease; LCI = lower 95% confidence interval; PAD = peripheral arterial disease; UCI upper 95% confidence interval

Figure S1. $PM_{2.5}$ annual average concentration in 2004 & 2016 using modeled estimates at 1x1 km resolution.



Annual $PM_{2.5}$ concentration in NC for the years 2004 (top) and 2016 (bottom).

Figure S2. $PM_{2.5}$ concentration-response curve generated using ground-based $PM_{2.5}$ monitoring network.