Supplement of:

**Associations of greenness, public parks and blue space with cardiovascular and respiratory disease hospitalization in the US Medicare cohort**

**Authors and affiliations:**

Jochem O Klompmaker a, b, Francine Laden a, b, c, Matthew H E M Browning d, Francesca Dominici e, S Scott Ogletree f, Alessandro Rigolon g, Jaime E Hart a, b\*, Peter James a, h\*

*a Department of Environmental Health, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, USA*

*b Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital, Boston, Massachusetts 02115, USA*

*c Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, USA*

*d Department of Parks, Recreation and Tourism Management, Clemson University, Clemson, South Carolina 29634, USA*

*e Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA*

*f OPENspace Research Centre, School of Architecture and Landscape Architecture, University of Edinburgh, Edinburgh, UK*

*g Department of City and Metropolitan Planning, The University of Utah, 375 South 1530 East, Salt Lake City, Utah 84112, USA*

*h Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts 02215, USA*

*\* Contributed equally as last author*

**Corresponding author:**

Jochem O Klompmaker, PhD

jklompmaker@hsph.harvard.edu

Department of Environmental Health, Harvard T. H. Chan School of Public Health, Landmark Center, 401 Park Drive, Boston Massachusetts 02215, USA

Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital, 181 Longwood Avenue, Boston, Massachusetts 02115, USA

**Table of contents:**

Supplemental Methods S1. Exposure assessment

Supplemental Methods S2. Covariates

Table S1. Descriptive statistics of the full and urban CVD and RSD cohorts (aggregated data)

Table S2. Results of sensitivity analyses

Figure S1. Spearman correlation in the full and urban CVD and RSD cohorts

Figure S2. Exposure-response curves with CVD and RSD hospitalization in the full and urban cohorts

Figure S3. Models with increasing adjustment for potential confounders in the full and urban cohorts

Figure S4. Results of effect modification by individual demographics and region in the urban cohort

Figure S5. Results of effect modification by neighborhood SES in the urban cohort

**Supplemental Methods**

**S1. Exposure assessment**

**- Public park cover**

We created a public park cover dataset based on the United States Geological Survey Protected Areas Database of the US (PAD-US). The PAD-US strives to be a complete inventory of public land and other protected areas in the US by compiling the “best available” data provided by land managing agencies and organizations.1 As PAD-US differentiates between multiple types of public lands, we selected all land types that are likely to be known and used by the general public for outdoor recreation from PAD-US V2.1 (2020) to create a public park cover dataset. This included only open and restricted access areas to provide an accessible and recreational version of the PAD-US ("PAD-US-AR"). To assess zip code-level public park cover, we converted the public park dataset to a raster image and calculated percent public park cover for each zip code.

The following designations were included in the public park cover dataset (PAD-US-AR):

1. Parks and open spaces open for public access or restricted access (i.e., seasonally open, fees required, or permits required) including but not limited to lands managed by the National Park Service, U.S. Forest Service, Bureau of Land Management, U.S. Fish & Wildlife, Army Corps of Engineers, State Parks, State Departments of Conservation, State Departments of Natural Resources, State Departments of Land, State Fish and Wildlife Departments, State Forest Service, State Park and Recreation Departments, Tennessee Valley Authority, and city and county park and recreation departments
2. Conservation easements

The following designations were excluded from the public park cover dataset (PAD-US-AR):

1. Department of Energy, Department of Defense, and Bureau of Reclamation lands
2. Marine areas managed as/by Marine Protected Areas, National Oceanic and Atmospheric Administration, Bureau of Ocean Energy Management, etc.
3. Proclamation areas, which are boundaries of national lands used for administrative purposes that overlap with large areas of public lands that are not all available to the public
4. Fish hatcheries and other lands used for water rights with regulated hunting
5. National Park easements (i.e., lands paralleling but not including the Appalachian Trail and not used by the public)
6. Joint management areas (i.e., university research stations)
7. Non-governmental organization lands (aside from conservation easements)
8. State trust/land survey lands
9. American Indian Lands
10. Other areas with unknown access or closed public access (i.e., limited to coordinated programs and research)

**- Greenness**

The Normalized Difference Vegetation Index (NDVI), an indicator of greenness, was estimated using satellite imagery. The NDVI is calculated as the ratio between the red and near infrared values.2 NDVI ranges from -1 to 1 and values close to 1 correspond to areas with complete coverage by live vegetation, values close to zero correspond to areas without live vegetation and negative values correspond to water. To assess greenness, we used images from Landsat 7 and Landsat 8 (Collection 1 Tier 1 DN values, representing scaled, calibrated at sensor radiance) for the entire US from June 1, up to August 31 (summer), for each year (2000-2016). Landsat 7 and Landsat 8 images are generated every 16 days at 30m resolution. Using Google Earth Engine,3 we made cloud-free composite maps for the contiguous US. After setting negative NDVI values to zero, we calculated the spatially weighted mean summer NDVI for each zip code in the US for each year.

**- Blue space cover**

We estimated blue space using satellite imagery based on the Joint Research Centre's Global Surface Water Dataset.4 This dataset contains information about the location and temporal distribution of surface water from 1984 to 2018, based on images from Landsat 5, 7, and 8 (30 m spatial resolution). To classify each pixel into water or non-water, we used the “Occurrence” band. This band indicates the frequency with which water was present in each pixel during the time period. We used a 50% threshold (water was present 50% or more of the time) to classify pixels into blue space and no blue space. As zip codes are used for postal services, adjacent water bodies such as rivers, lakes and oceans are not always included in zip code areas. Therefore, we calculated spatially weighted mean blue space of zip codes with a 1000 m buffer to be able to capture water bodies close to each zip code. We used Google Earth Engine 3 to calculate spatially weighted mean blue space for each zip code.

**S2. Covariates**

For each beneficiary, information about year of entry, age at year of Medicare entry, sex, race, Medicaid eligibility (a proxy for low socioeconomic status), and zip code of residence for all Medicare beneficiaries were derived from the Medicare beneficiary file. As a proxy for neighbourhood SES (nSES), we used eight zip code-level covariates from the US Census and American Community Survey (median home value, median household income, population density, percent Hispanic, percent Black, percent of the population with less than a high school degree, percent below the poverty level, and percent of owner-occupied housing units). We linked percent of the population that were ever smokers (based on the county level), derived from the nationwide Behavioural Risk Factor Surveillance System (BRFSS). BRFSS is the nation’s premier system of surveys that collect information about health-related risk behaviours of US residents. US census (2000, 2009-2016) and BRFSS (2000-2011) variables were available for some years but not all. Temporal interpolation using a moving average algorithm within each zip code was performed for missing years, as described previously.5

We also assessed zip code-level air pollution (particulate matter less than 2.5 microns (PM2.5) and nitrogen dioxide (NO2)), and meteorological indicators (maximum daily temperature, specific humidity, and precipitation). Briefly, annual PM2.5 and NO2 concentrations across the contiguous US for 2000-2016 were estimated based on predictions from well-validated spatio-temporal ensemble models.6–8 Detailed information about air pollution models can be found elsewhere.6–8 For each zip code, the annual average concentrations were estimated by averaging the estimations at grid cells whose centroids fall within the boundary of that ZIP code. Average summer maximum temperature, ambient specific humidity and total precipitation were estimated for each year for each zip code using daily data from the Gridded Surface Meteorological dataset.9

**Literature**

1 U.S. Geological Survey (USGS), Gap Analysis Project. Protected Areas Database of the United States (PAD-US) 2.1: U.S. Geological Survey data release. 2020. DOI:https://doi.org/10.5066/P92QM3NT.

2 NASA. Measuring Vegetation (NDVI & EVI). https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring\_vegetation\_2.php (accessed June 16, 2020).

3 Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens Environ* 2017; **202**: 18–27.

4 Pekel JF, Cottam A, Gorelick N, Belward AS. High-resolution mapping of global surface water and its long-term changes. *Nature* 2016; **540**: 418–22.

5 Di Q, Wang Y, Zanobetti A, *et al.* Air Pollution and Mortality in the Medicare Population. *N Engl J Med* 2017; **376**: 2513–22.

6 Requia WJ, Di Q, Silvern RF, *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**: 11037–47.

7 Di Q, Amini H, Shi L, *et al.* An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. *Environ Int* 2019; **130**: 104909.

8 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. *ACS Publ* 2019; **54**: 1372–84.

9 Abatzoglou JT. Development of gridded surface meteorological data for ecological applications and modelling. *Int J Climatol* 2013; **33**: 121–31.

**Table S1. Descriptive statistics of the cardiovascular disease (CVD) and respiratory disease (RSD) hospitalization cohorts 2000-2016 (aggregated data).a, b**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CVD hospitalization** | | **RSD hospitalization** | |
| **Covariate** | **Full cohort** | **Urban cohort** | **Full cohort** | **Urban cohort** |
| **Median (IQR)** | **Median (IQR)** | **Median (IQR)** | **Median (IQR)** |
| **Natural environment measures** |  |  |  |  |
| - % Public park cover | 7·5 (15·9) | 11·4 (13·8) | 7·5 (15·9) | 11·4 (13·8) |
| - NDVI | 0·52 (0·27) | 0·39 (0·23) | 0·52 (0·27) | 0·39 (0·23) |
| - % Blue space cover (1000m buffer) | 0·5 (3·2) | 0·7 (5·2) | 0·5 (3·2) | 0·7 (5·2) |
| - % Blue space cover (no buffer) | 0·3 (1·3) | 0·2 (1·3) | 0·3 (1·3) | 0·2 (1·3) |
| **US census covariates** |  |  |  |  |
| - Population density (persons/mile2) | 517·4 (2919·0) | 3602·0 (4269·1) | 511·7 (2890·8) | 3580·2 (4223·4) |
| - Median home value ($1,000) | 139·4 (145·3) | 198·0 (210·4) | 139·6 (145·0) | 198·0 (209·8) |
| - Median household income ($1,000) | 46·0 (24·9) | 51·5 (31·5) | 46·1 (24·9) | 51·7 (31·5) |
| - % with less than a high school degree | 24·7 (21·6) | 22·7 (21·1) | 24·5 (21·4) | 22·6 (20·9) |
| - % below the poverty level | 8·6 (8·2) | 8·6 (8·4) | 8·6 (8·2) | 8·6 (8·4) |
| - % owner-occupied housing units | 71·8 (21·4) | 60·1 (26) | 71·9 (21·3) | 60·1 (26·0) |
| - % Black | 3·7 (13·5) | 7·3 (17·9) | 3·7 (13·5) | 7·3 (18·0) |
| - % Hispanic | 5·0 (14·0) | 11·3 (22·8) | 5·0 (13·9) | 11·2 (22·6) |
| **BRFSS covariate** |  |  |  |  |
| - % ever smoked | 46·2 (9·1) | 44·3 (8) | 46·2 (9·1) | 44·3 (8·0) |
| **Other environmental exposures** |  |  |  |  |
| -summer temperature (°C) | 29·9 (5·2) | 29·6 (4·8) | 29·9 (5·2) | 29·6 (4·8) |
| -summer specific humidity (g of water vapor / kg of dry air) | 12·0 (4·0) | 11·7 (3·2) | 12·1 (4·0) | 11·7 (3·2) |
| -summer total precipitation (mm, daily total) | 3·1 (2·3) | 3·0 (3·2) | 3·1 (2·3) | 3·0 (3·1) |
| -PM2.5 (µg/m3) | 9·7 (4·0) | 10·2 (3·9) | 9·6 (3·9) | 10·1 (3·8) |
| -NO2 (ppb) | 16·3 (13·9) | 25·0 (13·0) | 16·1 (13·8) | 24·9 (13·0) |

a Descriptive statistics are given for the strata (aggregated data based on zip code, year, sex, race, Medicaid eligibility, 2-year categories of age at study entry and year of follow-up).

b Urban cohorts included all person years in zip codes with a population density of 1000+ persons/mile2.

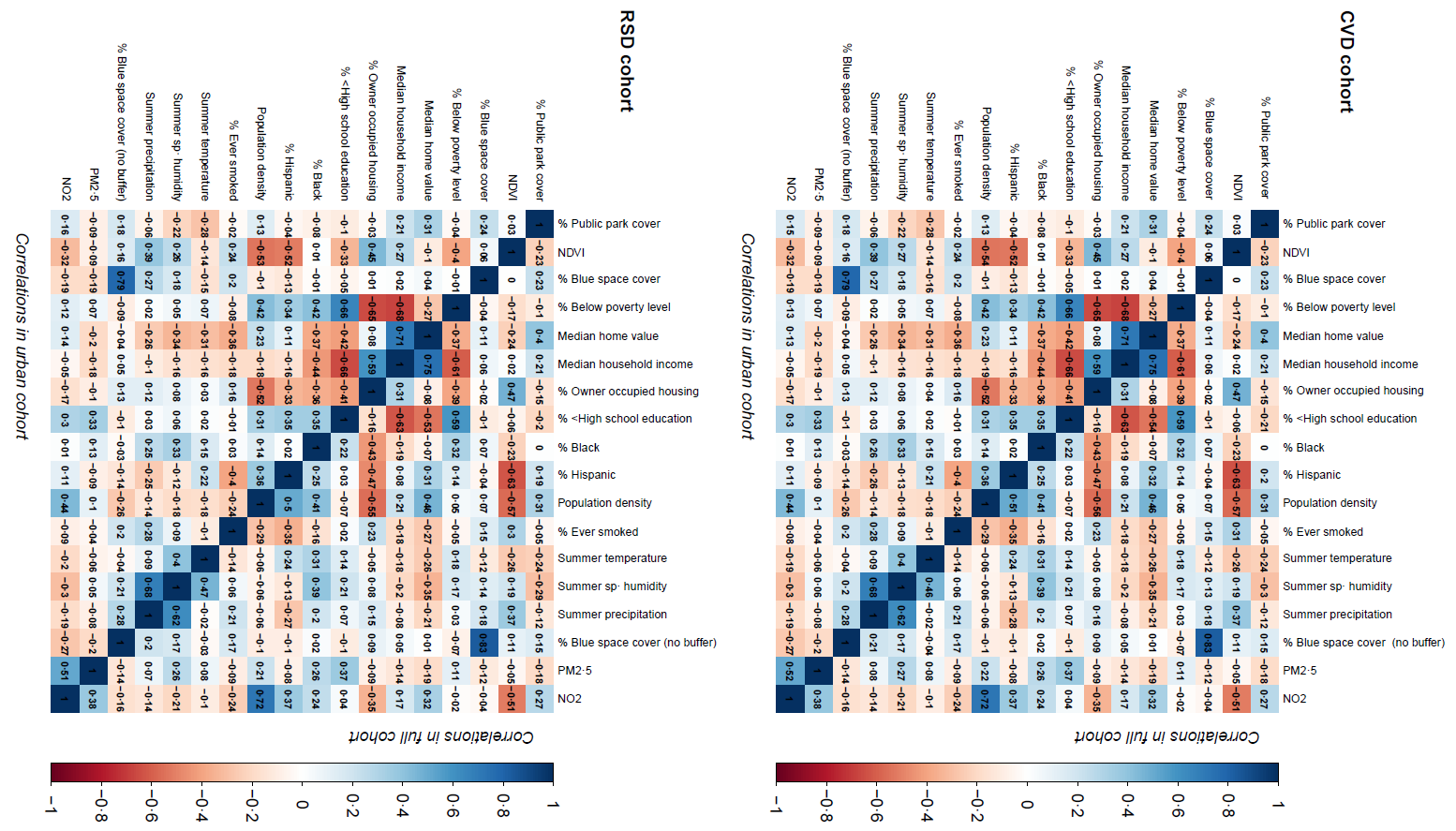
**­**

**Table S2. Sensitivity analyses of percent public park cover, NDVI and percent blue space cover with CVD and RSD hospitalization in the full and urban cohort of US Medicare fee-for-service beneficiaries aged ≥65 living in the contiguous US from 2000 through 2016.a, b**

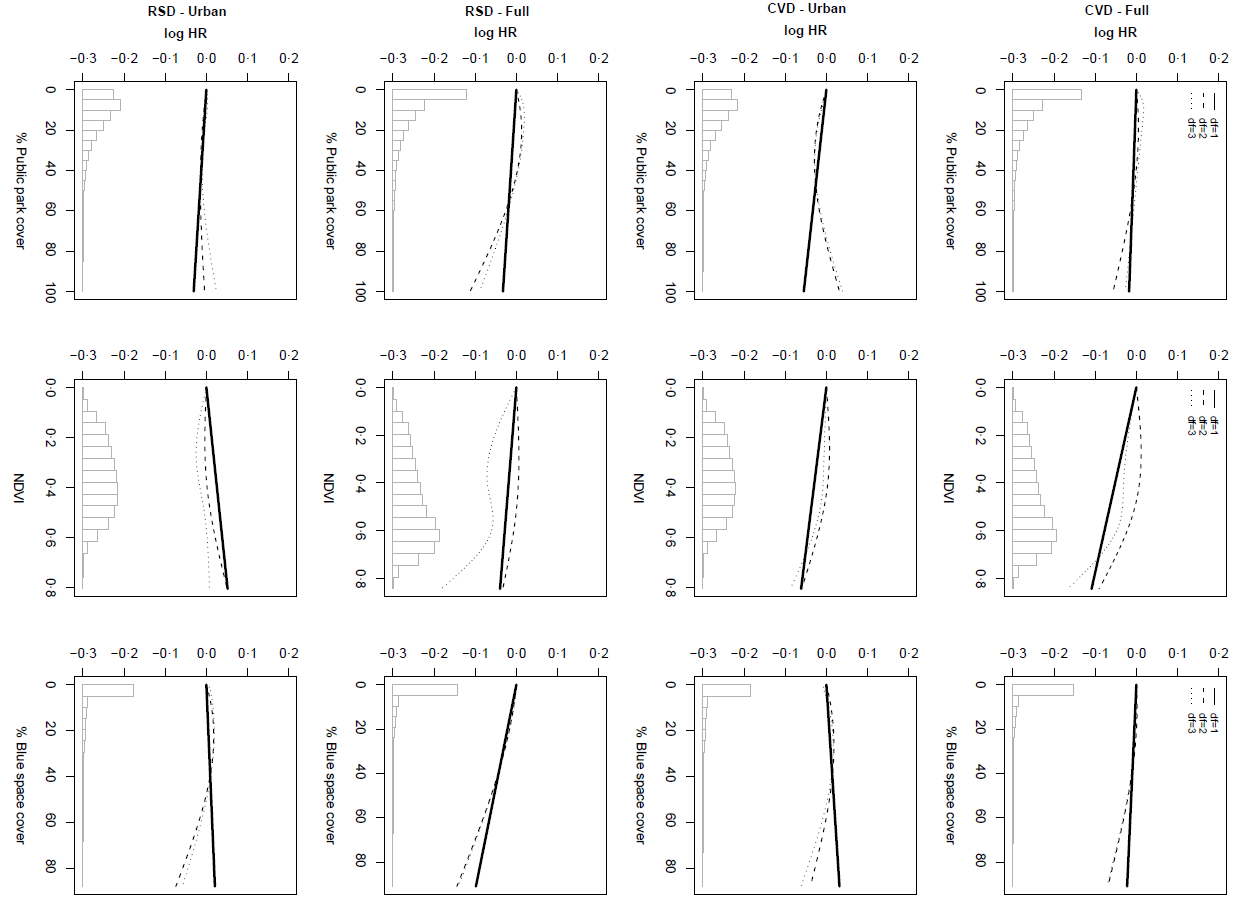
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Exposure (IQR)** | **main** | **Single- exposure** | **Adj. for PM2·5** | **Adj. for NO2** | **% Blue space cover (no buffer)** | **Excl. potentially prevalent cases** | **Death as competing risk** | |
| **CVD - Full** | | | | | | |
| **HR (95% CI)** | **HR (95% CI)** | **HR (95% CI)** | **HR (95% CI)** | **HR (95% CI)** | **HR (95% CI)** | **HR (95% CI)** | |
| **% Public park cover (15·9)** | 1·00 (0·99, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (0·99, 1·00) | 1·00 (0·99, 1·00) | 1·00 (0·99, 1·00) | 1·00 (0·99, 1·00) | |
| **NDVI (0·27)** | 0·97 (0·96, 0·97) | 0·97 (0·96, 0·97) | 0·96 (0·96, 0·97) | 0·99 (0·98, 0·99) | 0·96 (0·96, 0·97) | 0·96 (0·96, 0·97) | 0·96 (0·96, 0·97) | |
| **% Blue space cover (3·2)** | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | |
|  | **CVD – Urban** | | | | | | |
| **% Public park cover (15·9)** | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | |
| **NDVI (0·27)** | 0·98 (0·97, 0·99) | 0·98 (0·97, 0·98) | 0·98 (0·97, 0·99) | 0·99 (0·98, 1·00) | 0·98 (0·97, 0·99) | 0·98 (0·97, 0·99) | 0·97 (0·96, 0·98) | |
| **% Blue space cover (3·2)** | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | |
|  | **RSD – Full** | | | | | | |
| **% Public park cover (15·9)** | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | 1·00 (1·00, 1·01) | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | 1·00 (0·99, 1·00) | |
| **NDVI (0·27)** | 0·99 (0·98, 1·00) | 1·00 (0·99, 1·00) | 0·98 (0·98, 0·99) | 1·02 (1·01, 1·03) | 0·99 (0·98, 1·00) | 0·98 (0·97, 0·99) | 1·00 (0·99, 1·01) | |
| **% Blue space cover (3·2)** | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (0·99, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | |
|  | **RSD - Urban** | | | | | | | |
| **% Public park cover (15·9)** | 1·00 (0·99, 1·00) | 1·00 (0·99, 1·00) | 1·00 (0·99, 1·00) | 0·99 (0·99, 1·00) | 0·99 (0·99, 1·00) | 1·00 (0·99, 1·00) | 1·00 (0·99, 1·00) | |
| **NDVI (0·27)** | 1·02 (1·00, 1·03) | 1·02 (1·00, 1·03) | 1·02 (1·00, 1·03) | 1·03 (1·01, 1·04) | 1·02 (1·01, 1·03) | 1·01 (1·00, 1·03) | 1·02 (1·01, 1·04) | |
| **% Blue space cover (3·2)** | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·01) | 1·00 (1·00, 1·00) | 1·00 (1·00, 1·00) | |

a Associations are expressed per IQR increase of the CVD hospitalization (full) cohort. Models included percent public park cover, NDVI, percent blue space cover and were adjusted for calendar year, region, US census covariates, % ever smoked, summer temperature, summer specific humidity, summer total precipitation, an offset for total person-time and strata for all possible combinations of sex, race, Medicaid Eligibility, age at study entry (2-year categories), and follow-up year. Urban cohorts included all person years in zip codes with a population density of 1000+ persons/mile2.

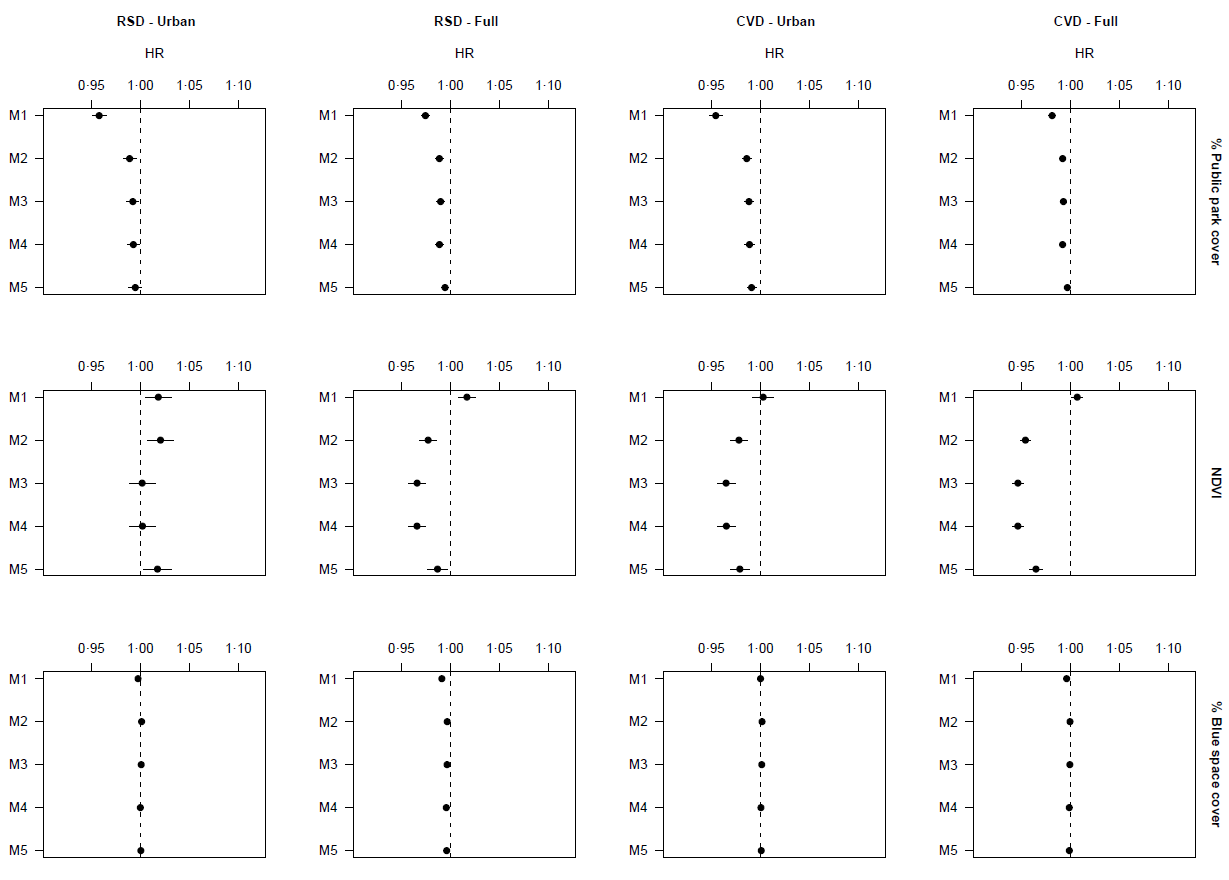
b Single-exposure = model including only a single exposure. Adj. for PM2·5 = main model additionally adjusted for PM2·5. Adj. for NO2 = main model additionally adjusted for NO2. % Blue space cover (no buffer) = model including blue space without a buffer. Excl. potentially prevalent cases = exclusion of individuals who had their first hospital admission within the first year of their follow-up and all records in the year 2000, Death as competing risk = inverse probability weights for dying were included as weights in the model. Inverse probability weights for dying were assessed by linking mortality data to our cohort. We fitted a model with death as outcome and included the same exposures and covariates as in our main model. Next, we calculated the inverse probability weight of dying for each strata and include the weights in our main model. Inverse probability weights lower than the 1st percentile were set to the 1st percentile and probabilities higher than the 99th percentile were set to the 99th percentile.

**Figure S1.** **Spearman correlations between percent public park cover, NDVI, percent blue space cover, SES covariates, smoking status, temperature, specific humidity, precipitation, PM2.5 and NO2.a**

a Correlations above the diagonal are for the full population, correlations below the diagonal are for the urban population

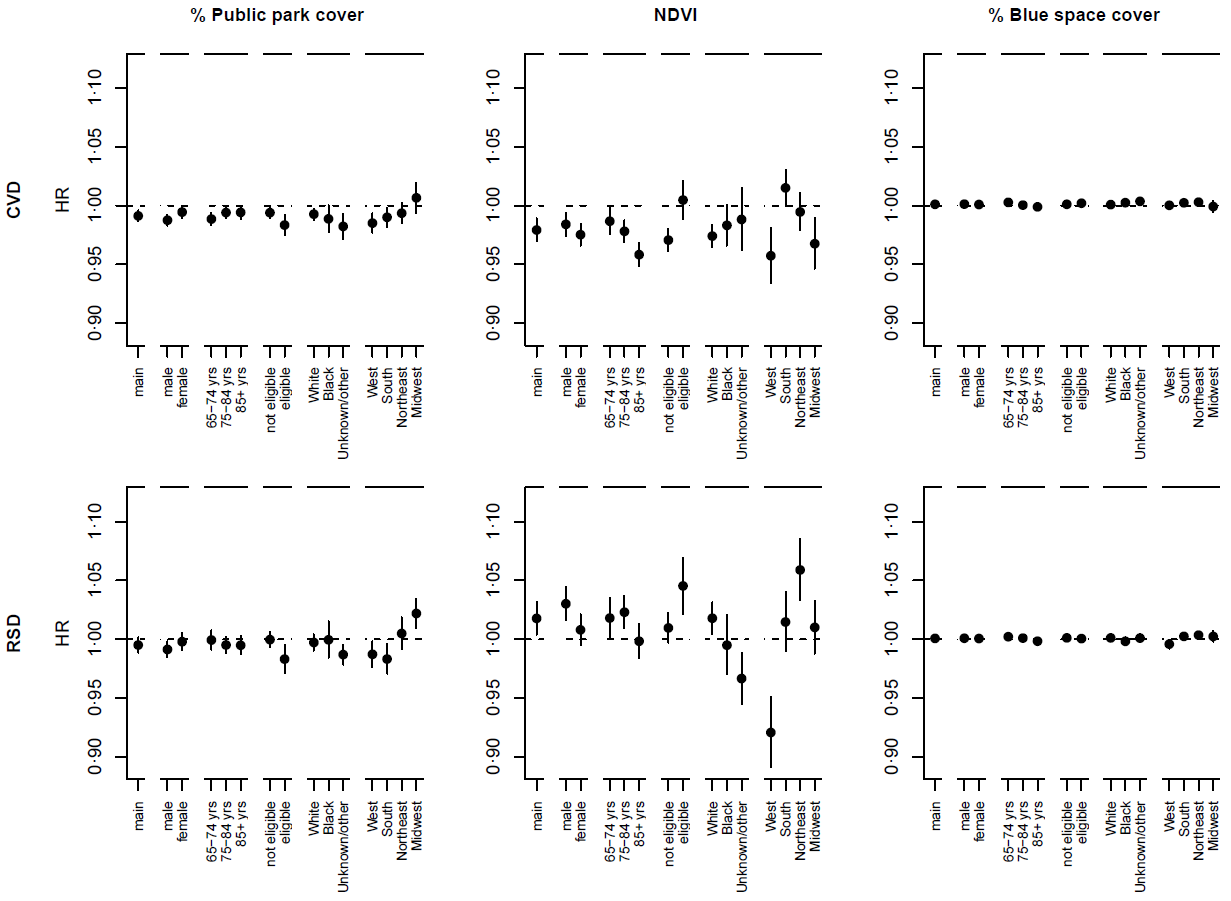
**Figure S2. Associations of percent public park cover, NDVI and percent blue space cover with CVD and RSD hospitalization in spline models with 1, 2 or 3 degrees of freedom in the full and urban cohort of US Medicare fee-for-service beneficiaries aged ≥65 living in the contiguous US from 2000 through 2016.a**

a Models included percent public park cover, NDVI, percent blue space cover and were adjusted for calendar year, region, US census covariates, % ever smoked, summer temperature, summer specific humidity and summer total precipitation, an offset for total person-time and strata for all possible combinations of sex, race, Medicaid Eligibility, age at study entry (2-year categories), and follow-up year. Urban cohorts included all person years in zip codes with a population density of 1000+ persons/mile2.



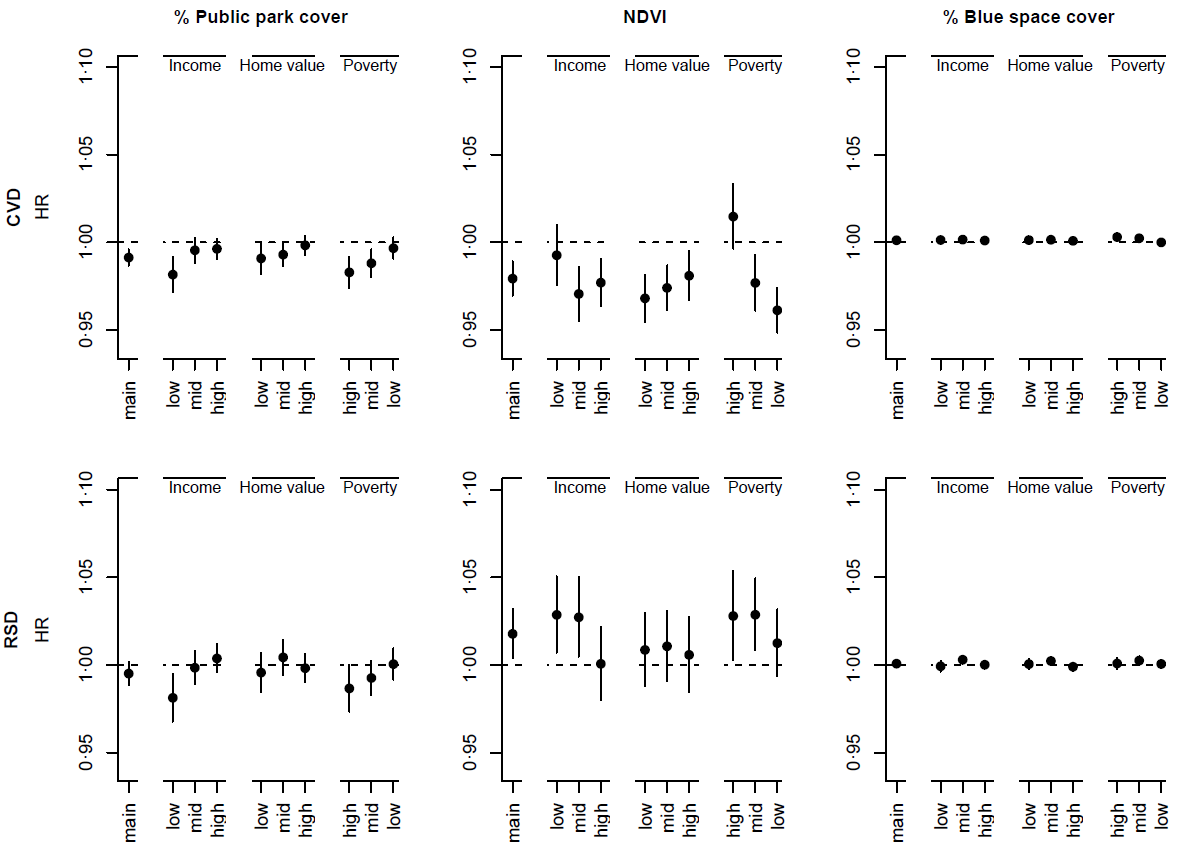
**Figure S3. Associations of percent public park cover, NDVI and percent blue space cover with CVD and RSD hospitalization in models with increasing adjustment for potential confounder in the full and urban cohort of US Medicare fee-for-service beneficiaries aged ≥65 living in the contiguous US from 2000 through 2016.a**

a Associations are expressed per IQR increase of the CVD hospitalization (full) cohort (IQR Percent public park cover: 15·9, NDVI: 0·27, Percent blue space cover: 3·2). Model 1 included percent public park cover, NDVI, percent blue space cover, calendar year, region and strata for all possible combinations of sex, race, dual, age at study entry (2-year categories) and follow-up year. In Model 2 all US census covariates (except population density) were added. In Model 3 population density was added. In Model 4 % ever smoked was added. In Model 5 summer temperature, specific humidity and total precipitation were added.



**Figure S4. Associations of percent public park cover, NDVI and percent blue space cover with CVD and RSD hospitalization in the urban population in stratified analyses by sex (male, female), age (65-74, 75-84, 85+ years), Medicaid eligibility (not eligible, eligible), race (White, Black, unknown/other) and region (West, South, Northeast, Midwest).a**

a Associations are expressed per IQR increase of the CVD hospitalization (full) cohort (IQR Percent public park cover: 15·9, NDVI: 0·27, Percent blue space cover: 3·2). Models included percent public park cover, NDVI, percent blue space cover and were adjusted for calendar year, region, US census covariates, % ever smoked, summer temperature, summer specific humidity and summer total precipitation, an offset for total person-time and strata for all possible combinations of sex, race, Medicaid Eligibility, age at study entry (2-year categories), and follow-up year.



**Figure S5. Associations of percent public park cover, NDVI and percent blue space cover with CVD and RSD hospitalization in the urban population in stratified analyses by median household income, median home value and percent below the poverty level.a, b**

a Associations are expressed per IQR increase of the CVD hospitalization (full) cohort (IQR Percent public park cover: 15·9, NDVI: 0·27, Percent blue space cover: 3·2). Models included public park cover, NDVI, percent blue space cover and were adjusted for calendar year, region, US census covariates, % ever smoked, summer temperature, summer specific humidity and summer total precipitation, an offset for total person-time and strata for all possible combinations of sex, race, Medicaid Eligibility, age at study entry (2-year categories), and follow-up year.

b To define strata, we used the following quantiles (q33.3, q66.7) for the CVD cohort: median household income ($1,000): 42.5, 62.4; median home value ($1,000): 146.8, 276.8; percent below the poverty level (%): 6.5, 11.7; for the RSD cohort: median household income ($1,000): 42.6, 62.5; median home value ($1,000): 147.0, 276.7; percent below the poverty level (%): 6.5, 11.6.