



# EPA Public Access

Author manuscript

*Environ Res.* Author manuscript; available in PMC 2021 April 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

*Environ Res.* 2020 April ; 183: 109176. doi:10.1016/j.envres.2020.109176.

## How do natural features in the residential environment influence women's self-reported general health? Results from cross-sectional analyses of a U.S. national cohort.

Wei-Lun Tsai<sup>a</sup>, Raquel A. Silva<sup>b,f</sup>, Maliha S. Nash<sup>c</sup>, Ferdouz V. Cochran<sup>b,g</sup>, Steven E. Prince<sup>a</sup>, Daniel J. Rosenbaum<sup>b</sup>, Aimee A. D'Aloisio<sup>d</sup>, Laura E. Jackson<sup>a</sup>, Megan H. Mehaffey<sup>a</sup>, Anne C. Neale<sup>a</sup>, Dale P. Sandler<sup>e</sup>, Timothy J. Buckley<sup>a,\*</sup>

<sup>a</sup>Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

<sup>b</sup>Oak Ridge Institute for Science and Education Research Fellow, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

<sup>c</sup>Office of Research and Development, U.S. Environmental Protection Agency, Newport, OR, USA

<sup>d</sup>Social & Scientific Systems, Durham, NC, USA

<sup>e</sup>National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

<sup>f</sup>Currently at ICF, Durham, NC, USA

<sup>g</sup>Currently at Carolinas Integrated Sciences & Assessments (CISA), Department of Geography, University of North Carolina at Chapel Hill, Chapel Hill, USA

### Abstract

**Background:** The relationship between health and human interaction with nature is complex. Here we conduct analyses to provide insights into potential health benefits related to residential proximity to nature.

**Objectives:** We aimed to examine associations between measures of residential nature and self-reported general health (SRGH), and to explore mediation roles of behavioral, social, and air quality factors, and variations in these relationships by urbanicity and regional climate.

**Methods:** Using residential addresses for 41,127 women from the Sister Study, a U.S.-based national cohort, we derived two nature exposure metrics, canopy and non-gray cover, using Percent Tree Canopy and Percent Developed Imperviousness from the National Land Cover Database. Residential circular buffers of 250m and 1250m were considered. Gradient boosted

\*Corresponding author. buckley.timothy@epa.gov (T.J. Buckley).

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

regression trees were used to model the effects of nature exposure on the odds of reporting better SRGH (Excellent/Very Good versus the referent, Good/Fair/Poor). Analyses stratified by urbanicity and regional climate (arid, continental, temperate) and mediation by physical activity, social support, and air quality was examined and were conducted.

**Results:** A 10% increase in canopy and non-gray cover within 1250m buffer was associated with 1.02 (95% CI: 1.00 – 1.03) and 1.03 (95% CI: 1.01 – 1.04) times the odds of reporting better SRGH, respectively. Stronger associations were observed for the urban group and for continental climate relative to other strata. Social support and physical activity played a more significant mediation role than air quality for the full study population.

**Discussion:** Findings from this study identified a small but important beneficial association between residential nature and general health. These findings could inform community planning and investments in neighborhood nature for targeted health improvements and potential societal and environmental co-benefits.

### Keywords

greenspace; ecosystem services; eco-health; climate; tree canopy; developed impervious; physical activity; social support; air quality

## 1. Introduction

A growing body of research recognizes that neighborhood natural environment is positively associated with an array of health benefits (Dadvand and Nieuwenhuijsen, 2019; Fong et al., 2018; Frumkin et al., 2017; James et al., 2015; Markevych et al., 2017; van den Bosch and Ode Sang, 2017) including physical activity (Astell-Burt et al., 2013; Sarkar et al., 2015), healthy weight (Tsai et al., 2019a; Villeneuve et al., 2018), mental health (Alcock et al., 2014; Banay et al., 2019; Beyer et al., 2014), and social cohesion (Jennings and Bamkole, 2019). Among the benefits, general health has also been positively associated with neighborhood natural environment (e.g., Dadvand et al., 2016; Reid et al., 2017); a common metric is self-reported general health (SRGH). This is a measure of an individual's perceived overall health status (Idler and Benyamini, 1997), which can be affected by many factors including health conditions, behaviors, and sociodemographic characteristics (Boerma et al., 2016; Gallagher et al., 2016; Heinze et al., 2015). SRGH has been shown to be reliable for evaluating individual health (Heistaro et al., 2001; Miilunpalo et al., 1997; Robert et al., 2014). For instance, a meta-analysis (DeSalvo et al., 2006) reported that people who reported "poor" health had 92% higher risk of all-cause mortality compared to those who reported "excellent" health. Additionally, SRGH has the advantage of capturing health more holistically and as perceived (Idler and Benyamini, 1997). Gender differences appear to exist in reported general health; evidence from fifty-nine countries indicated that women tended to report poorer health than men at all adult ages (Boerma et al., 2016).

The relationships between natural environment and human health can be influenced by humans' physical and conceptual interactions with the environment (Seymour, 2016). Biophilia theory describes the innate human need for connection with nature (Wilson, 1984).

Humans have been shown generally to perceive natural environment as healthy and possessing restorative features (Parsons, 1991; Ryan et al., 2014; Tidball, 2012).

Existing studies define natural environment in various ways, from a more restricted definition (e.g., a single type of natural landscape) to a broader inclusion of natural land cover (e.g., all landcover types except impervious). Many studies observed that tree cover was more strongly associated with better health outcomes (e.g., Akpinar et al., 2016 on mental health; Reid et al., 2017 on self-reported general health; Tsai et al., 2018 on frequent mental distress), while other studies indicated that not only greenery but also blue space (e.g., beach, lake) or gardens could provide health benefits (Brindley et al., 2018 on gardens; Gascon et al., 2017 on blue space; Volker and Kistemann, 2011 on blue space).

A study conducted in Barcelona, Spain, examined the relationships between residential natural environment and perceived general health using a broad measure of natural environment, the Normalized Difference Vegetation Index (NDVI), and both subjective and objective measures of park proximity (Dadvand et al., 2016). Their findings showed that NDVI greenness at 30m resolution and a subjective measure of park proximity were associated with increased odds of reporting better general health. Another study utilizing more finely-resolved land cover classes (3ft resolution) reported that only density of trees, not grass, was associated with increased odds of reporting very good or excellent health in New York City, United States (U.S.) (Reid et al., 2017). In a subsequent study, Reid et al. (2018) further investigated associations with perceived general health using NDVI at 30m, 250m, and 1km resolutions. However, the previously reported relationships were not observed using these measures. Studies are lacking, however, with respect to U.S. nationally representative data analyzed at the individual level (Silva et al., 2018).

It is important to understand the potential mechanisms underlying the relationships between natural environment and health. Many studies have suggested that the associations between nature and human health may be mediated by the provision of opportunities for physical activity and social interaction (Banay et al., 2019; Dadvand et al., 2016; Dzhambov et al., 2018; James et al., 2016; Richardson et al., 2013). For instance, green spaces provide heat mitigation and aesthetic settings that may encourage physical activity and social gatherings (Markevych et al., 2017). Several studies have considered the potential mechanism of air pollutant filtration by trees (Dzhambov et al., 2018; James et al., 2016; Orioli et al., 2019). Air pollutants are linked to many adverse health outcomes including preterm birth (Klepac et al., 2018), cardiovascular diseases (Dabass et al., 2018), cardiovascular mortality and all-cause mortality (Pope et al., 2018). Nowak et al. (2014) estimated that trees across the conterminous U.S. could remove 17.4 million tonnes of NO<sub>2</sub>, PM<sub>2.5</sub>, and O<sub>3</sub> annually, preventing approximately 850 air pollution-related deaths per year.

Relationships between natural environment and perceived general health may differ by sociodemographic factors and environmental settings. These variations are commonly addressed in statistical analyses by grouping participants using factors such as gender, age, socioeconomic factors (e.g., income, education), and urbanicity (Dadvand et al., 2016 for age and gender; Orioli et al., 2019 for age and gender; James et al., 2016 for urbanicity and income; Mass et al., 2006 for urbanicity and education). A few studies have examined

variation by other factors such as physical activity level (Banay et al., 2019) and climate zone (Jian et al., 2017). Disparate findings across environmental settings, such as urbanicity and climate zone, can be due to differences in abundance or exposure to natural resources. The provision of ecosystem services (e.g., heat mitigation, air pollutant filtration) (McDonnell et al., 2008; McDonnell and Pickett, 1990) and human perceptions about nature (Cleland et al., 2015; Soini et al., 2012) can also deviate across environmental settings.

Using data from the U.S.-based Nurses' Health Study cohort, James et al. (2016) observed stronger inverse associations between natural environment and mortality for participants in metropolitan areas than those in micropolitan and rural areas. Environmental settings can also be affected by climatic conditions. Jian et al. (2017) explored associations between environmental quality and mortality at the county level across the United States. They reported that, while the associations were not different by urbanicity, stronger inverse associations were found in dry regions in contrast to the other five climate zones (dry continental, hot summer continental, humid subtropical, Mediterranean, and warm summer continental). Associations may also differ across climate regions by type of measure. Tsai et al. (2019a) conducted residential address-level analyses and found that neighborhood street trees had consistent positive associations with healthy weight status in two U.S. cities with distinct climatic conditions--Portland, OR (temperate wet climate), and Phoenix, AZ (arid climate). Neighborhood streetside herbaceous cover was positively associated in the arid climate but negatively associated in the temperate climate with healthy weight status.

To address the need for nationally-representative data analyzed at the individual level, this study assesses associations between neighborhood natural environment and SRGH in women across the conterminous United States. We hypothesized that tree canopy and less impervious surfaces around the home are associated with better perceived general health, and that these associations are mediated by physical activity, social support, and air quality. In addition, we explored the variation in these relationships across urbanicity levels and climatic regions.

## 2. Methods

### 2.1 Health outcome and study population

SRGH data were obtained from the Sister Study (data release 6.0 with follow-up through 9/15/2016), a prospective cohort of 50,884 women across the U.S. and Puerto Rico (Sandler et al., 2017). The eligible participants were women aged 35 – 74 years old who did not have breast cancer at baseline but had at least one sister with a history of breast cancer (baseline: years 2003 to 2009). Data were collected for multiple domains, including demographic (e.g., age, race), socioeconomic (e.g., household income), behavioral (e.g., physical activity) and environmental (e.g., air pollution) factors, health conditions (e.g., body mass index, use of medications), and health and well-being outcomes (e.g., general health, quality of life). Participants have been contacted every two to three years since 2008 for detailed follow-up. To match with the year that their residential natural environment was assessed (i.e., 2011), SRGH at the first follow-up was used. This measure was collected by asking, “*In the past 24 months, would you say your health has generally been Excellent, Very Good, Good, Fair, or Poor?*” For this analysis, the responses were dichotomized as “Excellent/Very Good” (better

SRGH) versus the referent, “Good/Fair/Poor”, since this dichotomous classification has been shown to be an effective predictive health indicator in many studies (Maas et al., 2006; Reid et al., 2017; Reid et al., 2018; Simons, 2002).

Due to the available data coverage for natural environment, only participants who resided in the conterminous U.S. and whose residential locations were geocoded to street addresses were included for analysis (N=47,109) (Figure 1a). We further removed 177 participants who had less than 95% coverage of natural environment data within their analysis units. For example, participants residing close to the border between the U.S. and Canada (Figure 1b) were excluded. In addition, 5,805 participants who reported that they had moved between their baseline and first follow-up were removed to ensure that the outcome measure was collected when participants lived in the same residential locations as was registered at the study baseline. This additional removal resulted in a total number of 41,127 participants for analysis.

## 2.2 Measures of natural environment

To quantify various perspectives of nature, residential natural environment was estimated using tree cover and non-gray land cover (any land cover type other than impervious) based on satellite imaging data. Tree cover (canopy hereafter) derived from the U.S. Forest Service Tree Canopy Cartographic layer, part of the 2011 National Land Cover Database (NLCD) developed by the Multi-Resolution Land Characteristics Consortium (Homer et al., 2015). This layer provides information about the proportion of tree canopy within each 30m pixel; values range from 0 to 100. Non-gray land cover (non-gray hereafter) was derived from the 2011 NLCD Percent Developed Imperviousness map, which provides information about the proportion of impervious surface within each 30m pixel. Non-gray measures were calculated as 100 minus mean value of the 2011 NLCD Percent Developed Imperviousness within defined neighborhood extents. These two layers provide sufficient accuracy for estimating tree canopy and impervious cover (Homer et al., 2004). In part, our restriction to use of Sister Study first follow-up data (2008 – 2012) was justified to coincide with the latest data available to us from the U.S. Forest Service on percent tree canopy for 2011. For each residence, mean canopy and non-gray values were generated by delineating two circular analysis units with radii of 250m and 1250m. These buffers reflect a tight and broader aperture of neighborhood environments as has been employed by others (Banay et al., 2019; James et al., 2016). Considering the two sizes of buffer and two measures of natural environment, a total of four exposure measures per residence was available (Table 1).

## 2.3 Covariates

Covariates that may influence relationships between natural environment and SRGH involved socioeconomic and demographic factors, lifestyle variables, health conditions, and survey factors from the same survey period as the health outcome (Table 1). Socioeconomic and demographic factors at the individual level included age (continuous), household income (categorical: <20k as referent vs. 20 – 50k, 50 – 100k, 100 – 200k, and >200k), employment status (categorical: unable to work [i.e., medical leave, disability, and unemployment] as referent vs. working and unpaid/other occupation [i.e., retiree, homemaker, and student]), marital status (categorical: never as referent vs. married, divorced/separated, and widow),

educational attainment (categorical: high school or less as referent vs. associate college or some college, bachelor degree, and graduate degree), race/ethnicity (categorical: non-Hispanic White as referent vs. non-Hispanic Black, Hispanic, and other), and self-reported urbanicity (categorical: rural as referent vs. small town, suburban, and urban). In addition to individual-level consideration of socio-economic status, we also included census block level representation via the Area deprivation index (ADI). The ADI is from the Neighborhood Atlas (University of Wisconsin School of Medicine and Public Health, 2018); values are national percentiles ranging from 1 to 100 with higher values indicating more disadvantaged neighborhoods. The ADI is constructed from numerous indicators including educational distribution, household income, income disparity, occupational composition, unemployment rate, family poverty rate, single-parent household rate, home ownership, median home value, English language proficiency, and percentage of urban population (Singh, 2003). Inclusion of individual- and area- level covariates is consistent with prior studies (e.g., Dadvand et al., 2016; Reid et al., 2018) and justified since they are taking into account different aspects of socioeconomic characteristics.

Individual-level lifestyle variables included smoking status (categorical: never as referent vs. current and former), alcohol drinking status (categorical: never/former as referent vs. current social, and current regular), and dog ownership (binary: no as referent vs. yes). Health conditions included body mass index (continuous), menopausal status (binary: pre-menopause as referent vs. post-menopause), use of depression medication (binary: no as referent vs. yes), and use of anxiety medication (binary: no as referent vs. yes). Survey factors included survey season (categorical: winter as referent vs. spring, summer, and fall) and survey year (categorical: 2008 as referent vs. 2009, 2010, 2011, and 2012). All the covariates, except ADI, were collected from the Sister Study.

Factors involving variability of natural resources by environmental settings included urbanicity and climate zone. Urbanicity was based on individual responses from the Sister Study. We dichotomized the variable into two groups: non-urban (rural and small town) and urban (suburban and urban). The differences in climate were evaluated by the Köppen climate classification for the conterminous United States (Idaho State Climate Services, 1999). The Köppen climate classification includes five major zones worldwide: equatorial, arid, warm temperate, snow/continental, and polar climates (Kottek et al., 2006). In the conterminous U.S., polar climate is not included. We assigned a climate zone to each participant based on residential location. There was a very small portion of participants located in the equatorial climate zone (i.e., participants residing in the everglades and the keys of Florida, around 1.7%); we grouped these participants into the warm temperate climate zone. The classification is complex taking temperature and precipitation into account. The primary feature of arid climate is little precipitation. Continental climate is mainly characterized by temperature with at least one month averaging below  $-3^{\circ}\text{C}$  ( $26.6^{\circ}\text{F}$ ) and one month averaging above  $10^{\circ}\text{C}$  ( $50^{\circ}\text{F}$ ). For temperate climate the primary characteristic is that the coldest monthly temperature range between  $-3^{\circ}\text{C}$  ( $26.6^{\circ}\text{F}$ ) and  $18^{\circ}\text{C}$  ( $64.4^{\circ}\text{F}$ ) (Kottek et al., 2006).

## 2.4 Statistical analysis

The analytical approach for modeling the associations between SRGH and natural environment included logistic regression, stratified analysis, and mediation analysis (Figure 2). First, we built a crude model with controlling for commonly used covariates, including age, household income, employment status, educational attainment, race/ethnicity, and survey season. There is insufficient evidence that any of the variables available to us warrant a priori inclusion. Subsequently, we considered additional variables (i.e., lifestyle variables and health conditions mentioned in 2.3 Covariates) and constructed models gradient boosted regression trees (GBRT) for covariate selection. GBRT is a data-driven machine learning technique which builds models sequentially to improve prediction. The principle is to find the most appropriate loss function by trial and error from the learning processes and finalize a model with the minimal value of the defined loss function (Friedman, 2002; Natekin and Knoll, 2013). This approach has been widely and successfully used in many fields and represents an efficient, automated method for prediction and classification (e.g., Friedman and Meulman, 2003 in epidemiology; Moisen et al., 2006 in ecology; Naghibi et al., 2015 in geography). Covariates were included if their relative influence was greater than one. Twelve covariates were selected: age, household income, employment status, educational attainment, race/ethnicity, smoking status, drinking status, body mass index, use of depression medication, use of anxiety medication, survey season, and area deprivation index. Since similar effect sizes of natural environment were observed by these two model approaches with better concordance statistics (0.72) by the GBRT models (compared to the crude models (0.65), Supplemental Table S1). The final models were based on the GBRT approach.

The final logistic regression model was then applied to examine the relationships between SRGH and natural environment, controlling for the selected covariates. Multicollinearity was identified by the generalized variance inflation factors (GVIF). Covariates with GVIF values greater than three were considered to have multicollinearity. All the covariates had GVIF values of less than two, and therefore none of the covariates were removed. Effect modification was evaluated by inclusion of the multiplicative interaction term of each covariate and each measure of natural environment and testing whether it was significant ( $p < 0.05$ ) using likelihood ratio tests. We, however, did not observe a consistent effect modifier in this analysis.

Spatial autocorrelation was evaluated by Moran's  $I$  using model residuals with each measure of natural environment. Moran's  $I$  ranges from  $-1$  (spatially dispersed) to  $+1$  (spatial clustered) with a value of zero indicating spatial random distribution. All Moran's  $I$  values were less than 0.0002 for both approaches, indicating very weak spatial autocorrelation. Therefore, issues of spatial autocorrelation were not of concern in this study. We also employed natural cubic spline regression (four degrees of freedom) to detect potential non-linear relationships between natural environment and SRGH. Likelihood ratio tests showed no significant ( $p < 0.05$  for statistical significance) differences between models with and without spline terms, and therefore linear relationships were considered to be adequate for our model. All the statistical analyses were conducted using SAS version 9.4 (SAS Institute,

Cary NC, USA), except the selection for covariates based on GBRT approach which was conducted using ‘dismo’ package in RStudio with R version 3.4.3 (RStudio Team, 2015).

However, because there are fundamental differences in the natural environment with urbanicity and climate zone, we stratified our analysis by these variables. We also explored variations involving the six combinations of urbanicity and climate zone (urban in arid climate, non-urban in arid climate, urban in continental climate, non-urban in continental climate, urban in temperate climate, and non-urban in temperate climate).

## 2.5 Additional statistical analysis - Mediation analysis

Three categories of potential mediators between natural environment and SRGH were considered: behavioral, social, and air quality from the same survey period as the health outcome (Table 1). Behavioral mediators referred primarily to physical activity. The Sister Study provided information on weekly duration of several levels of physical activity (moderate physical activity, vigorous physical activity, and walking). Two measures of physical activity were used in this study: a combined measure of total hours for moderate and vigorous physical activity (MVPA) per week and total hours for walking per week (walking). We observed some implausible responses in total hours of physical activity measures, such as 255 activity hours per week. We carefully investigated outliers in MVPA and walking hours by determining if a participant reported total hours greater than the 95th percentile (18 hours for MVPA and 28 hours for walking). As a result, 2,626 and 1,910 responses (6.4% and 4.6%) were eliminated from mediation analysis by MVPA and walking, respectively.

Social mediators included support from family and friends. The Sister Study conducted a special survey for stress and coping at first detailed follow-up that collected variables related to social support such as, “*how often I can count on someone to provide me with emotional support?*”, “*how often I can count on someone if I need help?*”, “*how often there is someone in my immediate family who believes in me and wants me to succeed?*” and, “*how often there is someone in family makes me feel important or special?*”. The responses to each of these questions had five possible answers: none of the time, a little of the time, some of the time, most of the time, and all of the time. We scored the responses from 0 (none of the time) to 4 (all of the time). Sum of the scores for these four questions for each participant was used in the model. We included a second social-support variable (support network) based on the survey question, “*In general, how many relatives or friends do you feel close to?*” The responses were in five categories: none, 1 – 2, 3 – 5, 6 – 9, and 10 or more. We dichotomized the responses as less than six (referent) versus six or more.

Air quality mediators included average concentrations of particulate matter  $< 2.5 \mu\text{m}$  in aerodynamic diameter ( $\text{PM}_{2.5}$ ) and  $\text{NO}_2$  by address at year 2011. Modeled estimates of  $\text{PM}_{2.5}$  and  $\text{NO}_2$  concentrations in ambient air were derived based on different methods but both considered many factors, including land use/land cover, population density, road network characteristics, and distance to selected geographic features (Chan et al., 2015).

All the mediation analyses were performed using SAS (PROC CAUSALMED, version 9.4, SAS Institute, Cary NC, USA). This procedure decomposes the total effect into four



components based on the formulas developed by VanderWeele (2014). These four components are the effect that is due neither to interaction nor mediation (controlled direct effect), the effect that is due to interaction but not mediation (reference interaction), the effect that is due to both interaction and mediation (mediated interaction), and the effect that is due to mediation but not interaction (pure indirect effect). The procedure requires at least three types of variables: outcome, treatment, and mediator. To perform mediation analysis, we specified SRGH as the outcome variable, one factor described here at a time as the mediator variable, one measure of natural environment at a time as the treatment variable, and interaction between mediator and treatment variables. We additionally specified the selected covariates as the covariate variable for in this analysis. The procedure computes total, controlled direct, natural direct, and natural indirect effects. Percentages of natural indirect effects relative to the estimated total effect (i.e., the effect that are attributed to the mediator) were reported.

### 3. Results

#### 3.1 Self-reported general health and socioeconomics of participants and neighborhood

The reporting rates for SRGH status across measures of natural environment, selected covariates, mediators, and strata are detailed in Table 1. Most of the first follow-up surveys were collected in fall (45.9%). Mean age was 58.2 (sd = 8.8), and around 40% of the participants' reported household income was \$50,000 to \$100,000. Most participants were currently working (67.5%), had higher educational attainment (53.4% bachelor or graduate degree), and were non-Hispanic white (86.8%). The mean ADI was 35.3 (sd = 24.4), suggesting that great differences in socioeconomic status existed among participants' neighborhoods.

More than half of the participants reported never smoking, and almost 75% were current regular drinkers. Mean body mass index was in the overweight category (27.6 and sd of 6.1). About 20% and 14% of the participants reported having used depression or anxiety medication, respectively.

#### 3.2 Characteristics of natural environment, mediators, and groups for stratified analysis

Mean canopy values at 250m and 1250m neighborhood extents ("buffers") were 23.2% (sd = 20.8) and 24.4% (sd = 19.4) (Table 1). The non-gray measure had higher respective mean values of 72.5% (sd = 19.3) and 75.5% (sd = 17.8) at 250m and 1250m buffers. For any of the measures of natural environment, participants who reported "Excellent/Very Good" had higher mean values than those who reported "Good/Fair/Poor" (e.g., mean canopy at 250m for "Excellent/Very Good" group was 23.9% and 21.8% for "Good/Fair/Poor" group with  $t = 7.8$ ,  $p < 0.001$ ).

Potential mediators among participants are described in Table 1. Mean hours for weekly walking and MVPA were 4.2 (sd = 5.4) and 3.6 (sd = 4.1) hours, with higher mean values in the "Excellent/Very Good" group for both variables. Mean score for social support was 13.7 (sd = 2.8), with slightly higher mean score in the "Excellent/Very Good" group (14.1 vs. 13.0 for "Good/Fair/Poor" group). Around 55% of the participants reported having fewer

than six people that they felt comfortable with calling for help or sharing private matters. For those who reported having more than six people for their support network, the majority of participants reported better SRGH (73.4% in “Excellent/Very Good” group vs. 26.6% in “Good/Fair/Poor” group among the “more than six” group, not shown in Table 1).

For the characteristics among participants in different strata (Table 1 and Figure 3), around 60% of the participants resided in urban areas and more than half of the participants resided in the continental climate.

### 3.3 Relationships between SRGH and natural environment

A 10% increase in canopy at both 250m and 1250m buffer was associated with 1.02 times the odds of reporting “Excellent/Very Good” general health (95% CI: 1.01 – 1.03) (Table 2). Similar associations were observed at both buffer extents for each type of natural environment. The association for non-gray measures was larger than that observed for canopy. A 10% increase in the non-gray measure at both 250m and 1250m buffer was associated with 1.03 times the odds of reporting “Excellent/Very Good” general health (adjusted odds ratio [AOR]: 1.03, 95% CI: 1.01 – 1.04).

In urbanicity-stratified analyses (Table 3), we observed stronger associations between neighborhood natural environment and reporting better SRGH (Excellent/Very Good) only in the urban group for both buffers, except that the non-gray measure at 1250m was also associated with better SRGH in the non-urban group, although statistical evidence in differences in associations were not observed (respective p-values for interaction using canopy at 250, canopy at 1250m, non-gray at 250m, and non-gray at 1250m were 0.11, 0.46, 0.38, and 0.44). As expected, the mean values of natural environment were higher in the non-urban group than in the urban group. Associations were stronger at the closer neighborhood extent (250m) for participants who reported living in urban areas (e.g., AOR for a 10% increase in canopy at 250m: 1.03, 95% CI: 1.01 – 1.05 vs. 1.02, 95% CI: 1.00 – 1.04 at 1250m). Non-gray measures had similar associations as those observed using canopy at both neighborhood extents for the urban group.

In climate-stratified analyses, significant associations were observed only for the continental climate, although statistical evidence of differences in associations were also not observed (respective p-values for interaction using canopy at 250, canopy at 1250m, non-gray at 250m, and non-gray at 1250m were 0.09, 0.13, 0.06, and 0.27). Here, greater effects were observed at the smaller neighborhood extent with the non-gray measure. A 10% increase in canopy and in non-gray measure at 250m was both associated with 1.04 times the odds of reporting better SRGH (95% CI: 1.02 – 1.06 for both measures).

In stratified analyses with combinations of urbanicity and climate zone, for canopy, significant associations were observed only for the urban group in the continental climate. The effects of canopy in the urban group in the continental climate were at least 1.5 times greater than the effects observed separately in the urban group and in the continental climate. Statistical evidence for differences in associations was observed using canopy within 250m buffer (p-value for integration: 0.05) but not for canopy within 1250m buffer (p-value for interaction: 0.23). For non-gray measures, significant associations were

observed for both non-urban and urban groups in the continental climate at the 250m buffer; however, significant associations were observed only for the non-urban group in the continental climate at the 1250m buffer. There was no statistical evidence in the associations using non-gray measures (p-values for interaction using non-gray at 250m and 1250m: 0.06 and 0.38, respective). In the urban group in the continental climate, effect sizes for canopy were greater than those for non-gray measures (AOR for a 10% increase in canopy at 250m: 1.05, 95% CI: 1.02 – 1.08 vs. non-gray measure at 250m: 1.04, 95% CI: 1.02 – 1.07).

### 3.4 Mediation effects

MVPA was the only variable where a consistent mediation effect on the relationships between SRGH and natural environment was observed (Table 4). The effects were greater for non-gray measures than for canopy MVPA contributed 10.1% (95% CI: 0.1 – 20.0%) and 13.5% (95% CI: –0.9 – 27.9%) to the relationship between SRGH and canopy at 250m and 1250m, respectively; this variable also explained more than 19% of the relationships between SRGH and non-gray measures (19.6% with 95% CI: 7.9 – 31.3 at 250m and 23.6% with 95% CI: 6.85 – 40.3% at 1250m). Social support showed significant mediation effects only with non-gray measures. Around 15% of the relationship between SRGH and the non-gray measure at 250m and 1250m was explained by social support (respective 95% CI: 5.2 – 25.5% and 4.6 – 25.6%). No significant mediation effects were observed for walking or either of the air quality factors.

## 4. Discussion

A growing body of research supports that natural environment around residences is associated with better general health (Dadvand et al., 2016; de Vries et al., 2013; Reid et al., 2017). However, this relationship may be influenced by definitions of natural environments and methods of measurement, mediated by individual (behavioral), social, or air quality factors, and varied by environmental setting (urbanicity and climate zone). The current study examined the relationships between SRGH and residential natural environment among a U.S. national-scale population of women using a more restricted (canopy) and a broader (non-gray) definition of natural environment. This study is novel in being able to discern relatively subtle influences of natural-environment metrics and stratification variables in a large, national sample. We used a parsimonious model building approach via gradient boosted regression trees. Only variables that met our model building criteria were retained. Variations in the relationships between SRGH and natural environment were examined across divergent environmental settings by stratification for urbanicity (non-urban vs. urban) and climate zone (arid, continental, and temperate climates), where availability of and access to natural environment may greatly differ, and the roles of mediators were also explored.

### 4.1 Self-reported general health and neighborhood natural environment

Our findings corresponded to previous studies reporting that more tree cover (Reid et al., 2017) or greenness (Dadvand et al., 2016) was associated with better general health. Both canopy and non-gray measures were positively associated with increased odds of reporting “Excellent/Very Good” general health at the close (250m) and distant (1250m) neighborhood extents. Although the measures of greenness (commonly measured by NDVI)

and non-gray are not the same measure, they both provide information about the relative amount of overall natural environment within a study area. Reid et al. (2017) reported that tree cover but not grass cover was associated with better general health among residents in New York City. We, however, observed stronger associations using a non-gray measure (which includes everything except impervious surfaces) compared to only canopy among women across the entire United States. Using the same cohort as in our study, Villeneuve et al. (2018) reported a positive association between impervious surface and obesity. These findings suggest that aggregate landscape features other than gray cover may have positive effects on women's health, and inversely, that the presence of gray cover is associated with negative health effects.

Numerous studies call attention to methodological issues related to how natural environment is measured and how neighborhood extents are defined, as different methodologies may lead to disparate interpretations of the relationships between natural environment and health outcomes (e.g., Reid et al., 2018; Su et al., 2019; Tsai et al., 2019a). Using data with U.S. nationwide coverage, we observed similar effect sizes using close (250m) and distant (1250m) neighborhood extents for both measures of natural environment using 30m U.S. national-scale data. This observation is similar to the results of a study using 30m NDVI conducted in Barcelona, Spain (Dadvand et al., 2016). However, our findings differ from the study conducted in New York City using finely-resolved (3ft) land cover data (Reid et al., 2017), which reported stronger positive associations between tree cover and general health within larger (i.e., 1000m) buffers compared to more proximate (i.e., within 300m). James et al. (2016) found similar effect sizes while testing the effects of 250m NDVI on women's mortality across the U.S. with a similar design and neighborhood extents. The mixed findings across these studies may be due to large differences in the available underlying data, such as fine-resolution data do not have nationwide coverage. Given the potential for strong confounding and effect modification by many other spatially varying determinants of health, effect sizes may also differ between studies. These differences may also be caused by varying confounding control, as well as the distribution of effect modifiers between study samples.

#### 4.2 Variations across urbanicity and climate zone

Our results showed variations in the relationships between SRGH and natural environment by urbanicity and climate zone. Previous research (James et al., 2016) supports our findings that natural environment is more strongly associated with SRGH for participants who reside in urban areas than for others. For urban dwellers, AOR were slightly greater for smaller than larger buffers suggesting the more proximate natural features are more strongly associated with SRGH than the more distant features. The relationships between SRGH and natural environment varied across climates. Our findings differed from a previous study that examined associations between mortality and an environmental quality index (EQI) at the U.S. county level across six climatic zones (Jian et al., 2017). That study found stronger associations in the dry (arid) climate, whereas, in this study, relationships between SRGH and natural environment held only in the continental climate. One explanation is that the EQI and our natural environment measures likely do not represent the same types of environmental exposures; environmental measures in the EQI are predominantly indicators

of pollutant concentrations. Although significant associations were not found in our current study for arid and temperate climates, another study (Tsai et al., 2019a) that measured weight status as the health outcome reported significant protective effects of tree cover along walkable roads in both of these climates. Therefore, more research on health benefits conferred by natural environments across climate zones is warranted.

To our knowledge, this is the first study to examine variations in the relationships between human health and nature by both urbanicity and climate zone. Our findings revealed that the relationships might be subject to the type of natural environment. For canopy, associations varied by both urbanicity and climate zone, with stronger variations by urbanicity. Associations between SRGH and canopy held only for the urban group in the continental climate, and the effect sizes were almost three times greater than the effects observed for the full study population. With non-gray measures, variations depended mostly on climate zone. These findings suggest that distinctions in natural environment can be critical for landscape planning across hierarchies of environmental settings (e.g., urbanicity or combinations of urbanicity and climate zone).

### 4.3 Mediation roles of behavioral, social, and air quality factors

Our results indicated that MVPA and social support played mediating roles in the relationship between SRGH and exposures to natural environment. Mediation effects were more consistent with non-gray measures. The findings were partly consistent with a study that explored the mediation roles of social support, mental health, and physical activity on the relationships between general health and natural environment (Dadvand et al., 2016). That study found significant mediation effects by social support and mental health but not physical activity (an audit measure using the International Physical Activity Questionnaire, IPAQ). These different findings on the mediation role of physical activity may be due to differences in the survey instruments. The different findings could also be due to differences between the US and Spain in the ways people use natural environments. Similar mediation effects of physical activity have been observed specifically for mortality and obesity (James et al., 2016; Villeneuve et al., 2018). However, significant mediation effects from walking were not found in this study. It is possible that higher intensity physical activity provides additional health benefits (Nelson et al., 2007) that contribute to substantial differences in general health. The observed mediation by social support and physical activity also strengthens the inference that natural environment around homes has some associations with general health. While the natural environment signal is small, both of these mediators are salutogenic and have been demonstrated to respond positively to natural environment (Dadvand et al., 2016; James et al., 2017; Jennings and Bamkole, 2019; Vich et al., 2018).

We did not observe air quality factors (concentrations of PM<sub>2.5</sub> and NO<sub>2</sub> in ambient air) to have mediation effects for the full study population in this study, in contrast to other studies reporting that air pollution mediated associations between health outcomes and natural environment (e.g., James et al., 2016 on mortality; Orioli et al., 2019 on mortality and stroke incidence). It is unclear why we did not observe a mediating effect from air quality factors, but it could be due to the fact that we captured “self-reported” rather than objective health

outcomes or that our health metric is much more generally defined than those in the previous studies, or due to the differences in the way air pollution was measured.

#### 4.4 Limitations

The limitations of this study fall into three broad categories of study and statistical design, potential confounding not accounted, accuracy of the exposure measures. With respect to design, the fact that this study is cross-sectional and observational, limits our conclusion of influence to that of associative rather than causal. Specifically, we did not have information distinguishing between scenarios whereby areas with fewer trees or more gray space detracted from SRGH versus areas with fewer trees or more gray space in which individuals with worse SRGH reside or to which they move. An additional design limitation is the outcome measure of SRGH since it is perceived rather than objectively measured though this measure has been shown to be reliable for evaluating individual health (Heistaro et al., 2001; Miilunpalo et al., 1997; Robert et al., 2014). Although perceived health is not objectively measured, it can be argued that it represents the reality for that individual. The other design limitation relates to our study population being relatively homogeneous. Most of the participants are above age 55, non-Hispanic white, well-educated, and have a sister diagnosed with breast cancer. Additionally, because the cohort is all women, this may limit our ability to make inferences about the population at large.

There are likely a complex array of individual and area factors that influence the association between residential natural environment and SRGH. We were able to account for many but not all. First, we do not know the basis for self-selection of place of residence (Handy et al., 2005; Pinjari et al., 2009). Second, we were not able to account for neighborhood safety as a potentially confounding variable. (Kondo et al., 2017; Schusler et al., 2018). Similarly, population and housing density may be potentially confounding considerations that were not considered. However, urbanicity as a categorical covariate was considered but not selected into the final model by the GBRT approach.

We applied logistic regression to examine the associations between SRGH and natural environment. Evaluations of interactions in the context of logistic regression occur in log space and therefore caution in interpretation is warranted, such as they are not simply additive. Our analysis model (Figure 2) was structured to evaluate variations in the associations between natural environment and SRGH across urbanicity and climate zone strata. While sociodemographic factors, such as age, race/ethnicity, and neighborhood deprivation, may also lead to health disparities and inequity of the allocation of greenspaces (Jennings and Gaither, 2015; Schwarz et al., 2015), our analysis was limited to their inclusion as covariates rather than strata limiting their main effects analysis on the adjusted odds ratios. We examined only physical activity, social support, and air quality factors as potential mechanisms underlying the relationships between SRGH and natural environment. However, mental health may also contribute to these relationships. Many studies recognize that natural environment is associated with reduced stress (e.g., Ulrich et al., 1991) and improved mental health (e.g., de Vries et al., 2013), likely through psychological restoration (Kaplan and Kaplan, 1989). These restorative reactions can be elicited through human

perceptions by engaging with nature and thereby affect perceived health. This health pathway should be further explored.

Due to data availability, we could apply only medium resolution (30m) data to derive measures of natural environment, which may largely underestimate the amount of natural environment in urban areas (Nowak and Greenfield, 2010). In addition, configuration of natural environment (i.e., its pattern within the built environment) may be a stronger indicator than total amount of natural environment (Tsai et al., 2019b) and is recommended for use in research addressing relationships with human health (Termorshuizen and Opdam, 2009). We used only percentages of tree canopy and non-gray to examine relationships between SRGH and natural environment; these measures do not capture landscape complexity.

Finally, environmental exposures were examined only around individuals' residences within 1250m. This is a limitation because individuals are mobile and likely exposed to natural environments associated with other venues such as workplace and leisure destinations further from home (Hurvitz and Moudon, 2012; Moore et al., 2013). In addition, environmental exposures measured based on a fixed location may not accurately measure true amount (Hirsch et al., 2016; Holliday et al., 2017; Vich et al., 2018). Exposure buffers used in this study may also lead to the neighborhood effect averaging problem as described by Kwan (2018). This effect recognizes that environmental exposures estimated based on a fixed location (e.g., home, work) can over- or underestimate the true amount of exposures received from the actual activity routes of human mobility.

Despite several limitations in this study, the geographic distribution of study participants allowed us to examine natural environment-SRGH relationships at the U.S. national level and across environmental settings. In addition, multifaceted confounding factors were considered, and the results were similar for models constructed using different approaches. Potential mediators were also tested based on various factors. Our findings were consistent with other studies on SRGH and neighborhood natural environment for different populations and countries (Dadvand et al., 2016; de Vries et al., 2013; Reid et al., 2017). Findings from this study support the importance of nature around homes for general health status.

## 5. Conclusion

The relationship between human health and the environment is both important and complex. The focus of most environmental health research is on how human health is harmed by contaminants or other stressors in the environment. Here, we identify a small but important salutogenic association among features of the natural environment and general human health. Effects of natural environment on SRGH were protective and greater for non-gray measures relative to canopy, urban group relative to non-urban group, and continental climate relative to other climates. The magnitudes are consistent with other epidemiologic studies reporting associations with hazardous exposures (e.g., Bazzyar et al., 2019), but differ in identifying a salutogenic influence. Although the effect is small, it is of clear public health importance considering the approximately 80 million women within the relevant age range. Bastin et al. (2019) recently described the potential of forest restoration for climate

mitigation. Our findings relating to canopy cover especially near population centers provide additional motivation based on public health protection. Improved health can have profound economic benefits, given rising health care costs in the U.S., and city and regional planners may be well served to consider the potential societal value of protecting, conserving, and expanding environmental features associated with salutogenic influences.

The benefits of natural environment for human health have been widely recognized. Results from this study are generally consistent with and contribute to published literature related to the influence of the natural environment on human health. We build on the existing literature by identifying the influences of climatic characteristics as well as wide ranging mediation effects across social, environmental, and behavioral domains. We are building on this research using more finely-resolved measures of natural environment (higher resolution data and pattern metrics such as street-level natural environment) to better capture small patches of natural environment embedded in urban areas. We also plan to acquire additional human health data and apply experimental designs (e.g., before vs. after re-location) and longitudinal approaches, particularly to individuals who moved. These efforts will increase our understanding of the influences of landscape complexity and the accessibility of ecosystem services on human health.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## 6. References

- Akpinar A, et al., 2016 Does green space matter? Exploring relationships between green space type and health indicators. *Urban Forestry & Urban Greening*. 20, 407–418.
- Alcock I, et al., 2014 Longitudinal effects on mental health of moving to greener and less green urban areas. *Environmental Science & Technology*. 48, 1247–1255. [PubMed: 24320055]
- Astell-Burt T, et al., 2013 Mental health benefits of neighbourhood green space are stronger among physically active adults in middle-to-older age: Evidence from 260,061 Australians. *Preventive Medicine*. 57, 601–606. [PubMed: 23994648]
- Banay RF, et al., 2019 Greenness and depression incidence among older women. *Environmental Health Perspectives*. 127, 027001.
- Bastin J-F, et al., 2019 The global tree restoration potential. *Science*. 365, 76–79. [PubMed: 31273120]
- Bazyar J, et al., 2019 A comprehensive evaluation of the association between ambient air pollution and adverse health outcomes of major organ systems: a systematic review with a worldwide approach. *Environ Sci Pollut Res Int*. 26, 12648–12661. [PubMed: 30903465]
- Beyer K, et al., 2014 Exposure to Neighborhood green space and mental health: Evidence from the Survey of the Health of Wisconsin. *International Journal of Environmental Research and Public Health*. 11, 3453. [PubMed: 24662966]
- Boerma T, et al., 2016 A global assessment of the gender gap in self-reported health with survey data from 59 countries. *BMC public health*. 16, 675–675. [PubMed: 27475755]
- Brindley P, et al., 2018 Domestic gardens and self-reported health: a national population study. *International Journal of Health Geographics*. 17, 31. [PubMed: 30064435]
- Chan SH, et al., 2015 Long-term air pollution exposure and blood pressure in the Sister Study. *Environ Health Perspect*. 123, 951–8. [PubMed: 25748169]
- Cleland V, et al., 2015 Associations between the perceived environment and physical activity among adults aged 55–65 years: does urban-rural area of residence matter? *J Aging Phys Act*. 23, 55–63. [PubMed: 24412944]



- Dabass A, et al., 2018 Systemic inflammatory markers associated with cardiovascular disease and acute and chronic exposure to fine particulate matter air pollution (PM2.5) among US NHANES adults with metabolic syndrome. *Environmental Research*. 161, 485–491. [PubMed: 29223110]
- Dadvand P, et al., 2016 Green spaces and general health: Roles of mental health status, social support, and physical activity. *Environment International*. 91, 161–167. [PubMed: 26949869]
- Dadvand P, Nieuwenhuijsen M, Green space and health Integrating Human Health into Urban and Transport Planning. Springer, 2019, pp. 409–423.
- de Vries S, et al., 2013 Streetscape greenery and health: stress, social cohesion and physical activity as mediators. *Soc Sci Med*. 94.
- DeSalvo KB, et al., 2006 Mortality prediction with a single general self-rated health question. A meta-analysis. *Journal of general internal medicine*. 21, 267–275. [PubMed: 16336622]
- Dzhambov A, et al., 2018 Urban residential greenspace and mental health in youth: Different approaches to testing multiple pathways yield different conclusions. *Environmental Research*. 160, 47–59. [PubMed: 28961469]
- Fong KC, et al., 2018 A review of epidemiologic studies on greenness and health: Updated literature through 2017. *Current environmental health reports*. 5, 77–87. [PubMed: 29392643]
- Friedman JH, 2002 Stochastic gradient boosting. *Computational Statistics & Data Analysis*. 38, 367–378.
- Friedman JH, Meulman JJ, 2003 Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*. 22, 1365–1381. [PubMed: 12704603]
- Frumkin H, et al., 2017 Nature contact and human health: A research agenda. *Environmental Health Perspectives*. 125, 075001–075001. [PubMed: 28796634]
- Gallagher JE, et al., 2016 Factors associated with self-reported health: implications for screening level community-based health and environmental studies. *BMC Public Health*. 16, 640. [PubMed: 27460934]
- Gascon M, et al., 2017 Outdoor blue spaces, human health and well-being: A systematic review of quantitative studies. *International Journal of Hygiene and Environmental Health*. 220, 1207–1221. [PubMed: 28843736]
- Handy S, et al., 2005 Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transp Res Part D: Transp Environ*. 10.
- Heinze JE, et al., 2015 Relationships among disease, social support, and perceived health: A lifespan approach. *American Journal of Community Psychology*. 56, 268–279. [PubMed: 26462808]
- Heistaro S, et al., 2001 Self rated health and mortality: a long term prospective study in eastern Finland. *Journal of Epidemiology and Community Health*. 55, 227–232. [PubMed: 11238576]
- Hirsch JA, et al., 2016 Destinations that older adults experience within their GPS activity spaces: Relation to objectively measured physical activity. *Environment and Behavior*. 48, 55–77. [PubMed: 26783370]
- Holliday KM, et al., 2017 Are buffers around home representative of physical activity spaces among adults? *Health & Place*. 45, 181–188. [PubMed: 28391129]
- Homer C, et al., 2015 Completion of the 2011 National Land Cover Database for the conterminous United States—representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*. 81, 345–354.
- Homer C, et al., 2004 Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering and Remote Sensing*. 70, 82–840.
- Hurvitz PM, Moudon AV, 2012 Home versus nonhome neighborhood: Quantifying differences in exposure to the built environment. *American Journal of Preventive Medicine*. 42, 411–417. [PubMed: 22424255]
- Idaho State Climate Services, 1999 Koppen climate classification for the conterminous United States. Available at <http://cloud.insideidaho.org/webApps/util/zip.ashx?fn=http://cloud.insideidaho.org/data/ics/koppen>.
- Idler EL, Benyamini Y, 1997 Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav*. 38, 21–37. [PubMed: 9097506]

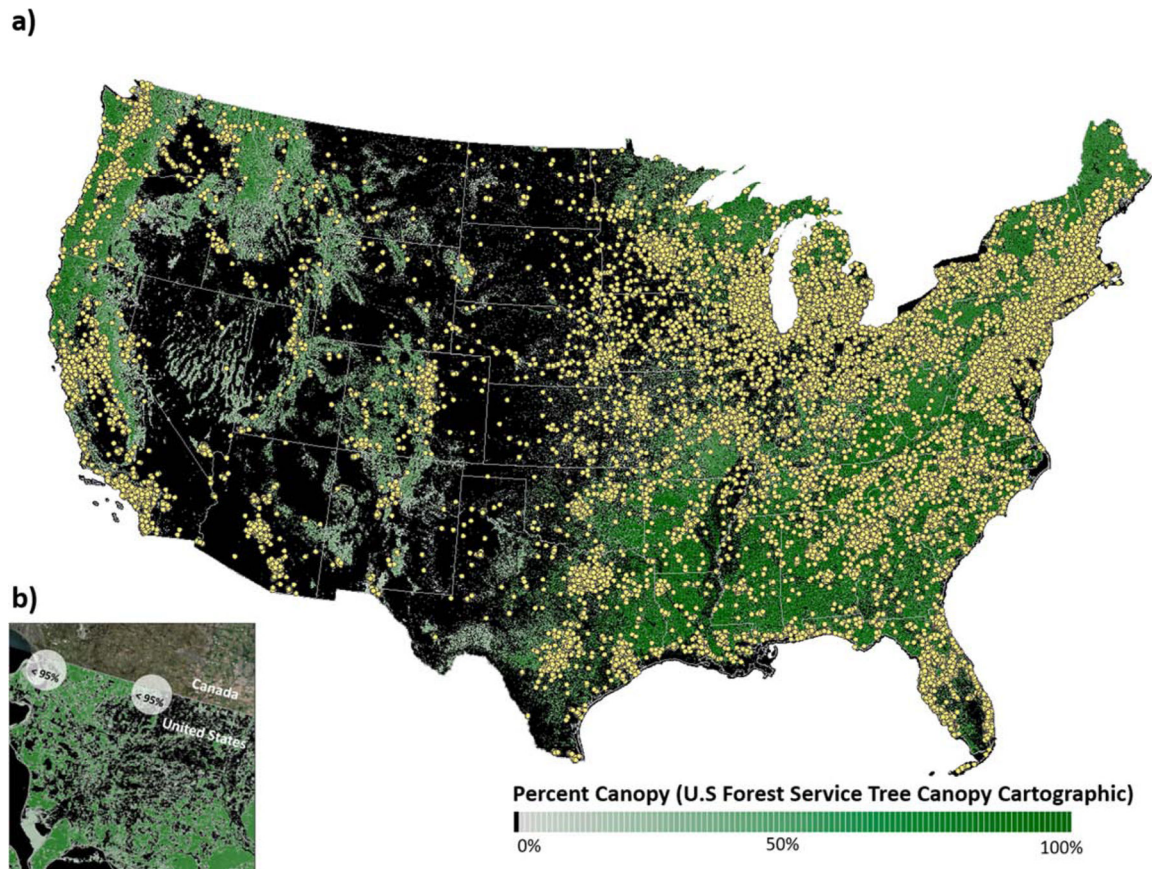
- James P, et al., 2015 A review of the health benefits of greenness. *Current Epidemiology Reports*. 2, 131–142. [PubMed: 26185745]
- James P, et al., 2016 Exposure to greenness and mortality in a nationwide prospective cohort study of women. *Environ Health Perspect*. 124, 1344–52. [PubMed: 27074702]
- James P, et al., 2017 GPS-based exposure to greenness and walkability and accelerometry-based physical activity. *Cancer Epidemiology Biomarkers & Prevention*. 26, 525–532.
- Jennings V, Bamkole O, 2019 The relationship between social cohesion and urban green space: An avenue for health promotion. *International Journal of Environmental Research and Public Health*. 16, 452.
- Jennings V, Gaither C, 2015 Approaching environmental health disparities and green spaces: An ecosystem services perspective. *International Journal of Environmental Research and Public Health*. 12, 1952. [PubMed: 25674782]
- Jian Y, et al., 2017 Associations between environmental quality and mortality in the contiguous United States, 2000–2005. *Environ Health Perspect*. 125, 355–362. [PubMed: 27713110]
- Kaplan R, Kaplan S, 1989 The experience of nature: A psychological perspective. CUP Archive.
- Klepac P, et al., 2018 Ambient air pollution and pregnancy outcomes: A comprehensive review and identification of environmental public health challenges. *Environmental Research*. 167, 144–159. [PubMed: 30014896]
- Kondo MC, et al., 2017 The association between urban trees and crime: Evidence from the spread of the emerald ash borer in Cincinnati. *Landscape and Urban Planning*. 157, 193–199.
- Kottek M, et al., 2006 World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*. 15, 259–263.
- Kwan M-P, 2018 The Neighborhood Effect Averaging Problem (NEAP): An elusive confounder of the neighborhood effect. *International Journal of Environmental Research and Public Health*. 15, 1841.
- Maas J, et al., 2006 Green space, urbanity, and health: how strong is the relation? *Journal of epidemiology and community health*. 60, 587–592. [PubMed: 16790830]
- Markevych I, et al., 2017 Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environ Res*. 158, 301–317. [PubMed: 28672128]
- McDonnell MJ, et al., Ecosystem processes along an urban-to-rural gradient *Urban Ecology*. Springer, 2008, pp. 299–313.
- McDonnell MJ, Pickett STA, 1990 Ecosystem structure and function along urban-rural gradients: An unexploited opportunity for ecology. *Ecology*. 71, 1232–1237.
- Miilunpalo S, et al., 1997 Self-rated health status as a health measure: The predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *Journal of Clinical Epidemiology*. 50, 517–528. [PubMed: 9180644]
- Moisen GG, et al., 2006 Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling*. 199, 176–187.
- Moore K, et al., 2013 Home and work neighbourhood environments in relation to body mass index: the Multi-Ethnic Study of Atherosclerosis (MESA). *Journal of Epidemiology and Community Health*. 67, 846–853. [PubMed: 23868527]
- Naghbi SA, et al., 2015 GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environmental Monitoring and Assessment*. 188, 44. [PubMed: 26687087]
- Natekin A, Knoll A, 2013 Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*. 7, 21–21. [PubMed: 24409142]
- Nelson ME, et al., 2007 Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. *Circulation*. 116, 1094. [PubMed: 17671236]
- Nowak DJ, Greenfield EJ, 2010 Evaluating the National Land Cover Database tree canopy and impervious cover estimates across the conterminous United States: a comparison with photo-interpreted estimates. *Environmental management*. 46, 378–390. [PubMed: 20676888]

- Nowak DJ, et al., 2014 Tree and forest effects on air quality and human health in the United States. *Environmental Pollution*. 193, 119–129. [PubMed: 25016465]
- Orioli R, et al., 2019 Exposure to residential greenness as a predictor of cause-specific mortality and stroke incidence in the Rome Longitudinal Study. *Environmental Health Perspectives*. 127, 027002.
- Parsons R, 1991 The potential influences of environmental perception on human health. *Journal of Environmental Psychology*. 11, 1–23.
- Pinjari AR, et al., 2009 Residential self-selection effects in an activity time-use behavior model. *Trans Res Part B: Met*. 43.
- Pope CA, et al., 2018 Mortality risk and PM2.5 air pollution in the USA: An analysis of a national prospective cohort. *Air Quality, Atmosphere & Health*. 1–8.
- Reid C, et al., 2017 Is all urban green space the same? A comparison of the health benefits of trees and grass in New York City. *International Journal of Environmental Research and Public Health*. 14, 1411.
- Reid CE, et al., 2018 It's not easy assessing greenness: A comparison of NDVI datasets and neighborhood types and their associations with self-rated health in New York City. *Health & place*. 54, 92–101. [PubMed: 30248597]
- Richardson EA, et al., 2013 Role of physical activity in the relationship between urban green space and health. *Public Health*. 127, 318–324. [PubMed: 23587672]
- Robert B, et al., 2014 Self-rated Health and medical outcomes in the Women's Health Initiative: The Aging Continuum, Health, Morbidity, Mortality. 3, 139.
- RStudio Team, 2015 RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.
- Ryan CO, et al., 2014 Biophilic design patterns: Emerging nature-based parameters for health and well-being in the built environment. 2014 8, 15.
- Sandler DP, et al., 2017 The Sister Study cohort: Baseline methods and participant characteristics. *Environ Health Perspect*. 125, 127003. [PubMed: 29373861]
- Sarkar C, et al., 2015 Exploring associations between urban green, street design and walking: Results from the Greater London boroughs. *Landscape and Urban Planning*. 143, 112–125.
- Schusler T, et al., 2018 Research note: Examining the association between tree canopy, parks and crime in Chicago. *Landscape and Urban Planning*. 170, 309–313.
- Schwarz K, et al., 2015 Trees grow on money: Urban tree canopy cover and environmental justice. *PloS one*. 10, e0122051. [PubMed: 25830303]
- Seymour V, 2016 The human-nature relationship and its impact on health: A critical review. *Frontiers in public health*. 4, 260–260. [PubMed: 27917378]
- Silva RA, et al., Advancing environmental epidemiology to assess the beneficial influence of the natural environment on human health and well-being. ACS Publications, 2018.
- Simons J, 2002 How is your health in general. Qualitative and quantitative studies on self-assessed health and socioeconomic differences herein.
- Singh GK, 2003 Area deprivation and widening inequalities in US mortality, 1969–1998. *Am J Public Health*. 93, 1137–43. [PubMed: 12835199]
- Soini K, et al., 2012 Residents' sense of place and landscape perceptions at the rural–urban interface. *Landscape and Urban Planning*. 104, 124–134.
- Su JG, et al., 2019 Associations of green space metrics with health and behavior outcomes at different buffer sizes and remote sensing sensor resolutions. *Environment International*. 126, 162–170. [PubMed: 30798197]
- Termorshuizen JW, Opdam P, 2009 Landscape services as a bridge between landscape ecology and sustainable development. *Landscape Ecology*. 24, 1037–1052.
- Tidball KG, 2012 Urgent biophilia: Human-nature interactions and biological attractions in disaster resilience. *Ecology and Society*. 17.
- Tsai W-L, et al., 2019a Associations between types of greenery along neighborhood roads and weight status in different climates. *Urban Forestry & Urban Greening*. 41, 104–117. [PubMed: 31031577]

- Tsai W-L, et al., 2018 Relationships between characteristics of urban green land cover and mental health in U.S. metropolitan areas. *International Journal of Environmental Research and Public Health*. 15, 340.
- Tsai W-L, et al., 2019b Street-level neighborhood greenery linked to active transportation: A case study in Milwaukee and Green Bay, WI, USA. *Landscape and Urban Planning*. 191, 103619.
- Ulrich RS, et al., 1991 Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*. 11, 201–230.
- University of Wisconsin School of Medicine and Public Health, 2018 Area Deprivation Index. Available at: <https://www.neighborhoodatlas.medicine.wisc.edu/>.
- van den Bosch M, Ode Sang Å, 2017 Urban natural environments as nature-based solutions for improved public health – A systematic review of reviews. *Environmental Research*. 158, 373–384. [PubMed: 28686952]
- VanderWeele TJ, 2014 A unification of mediation and interaction: a 4-way decomposition. *Epidemiology (Cambridge, Mass.)*. 25, 749–761.
- Vich G, et al., 2018 Green exposure of walking routes and residential areas using smartphone tracking data and GIS in a Mediterranean city. *Urban Forestry & Urban Greening*.
- Villeneuve PJ, et al., 2018 Association of residential greenness with obesity and physical activity in a US cohort of women. *Environ Res*. 160, 372–384. [PubMed: 29059619]
- Volker S, Kistemann T, 2011 The impact of blue space on human health and well-being - Salutogenetic health effects of inland surface waters: A review. *International Journal of Hygiene and Environmental Health*. 214, 449–460. [PubMed: 21665536]
- Wilson EO, *Biophilia*. Harvard University Press, Massachusetts, 1984.
- Zhang J, Yu KF, 1998 What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Jama*. 280, 1690–1. [PubMed: 9832001]

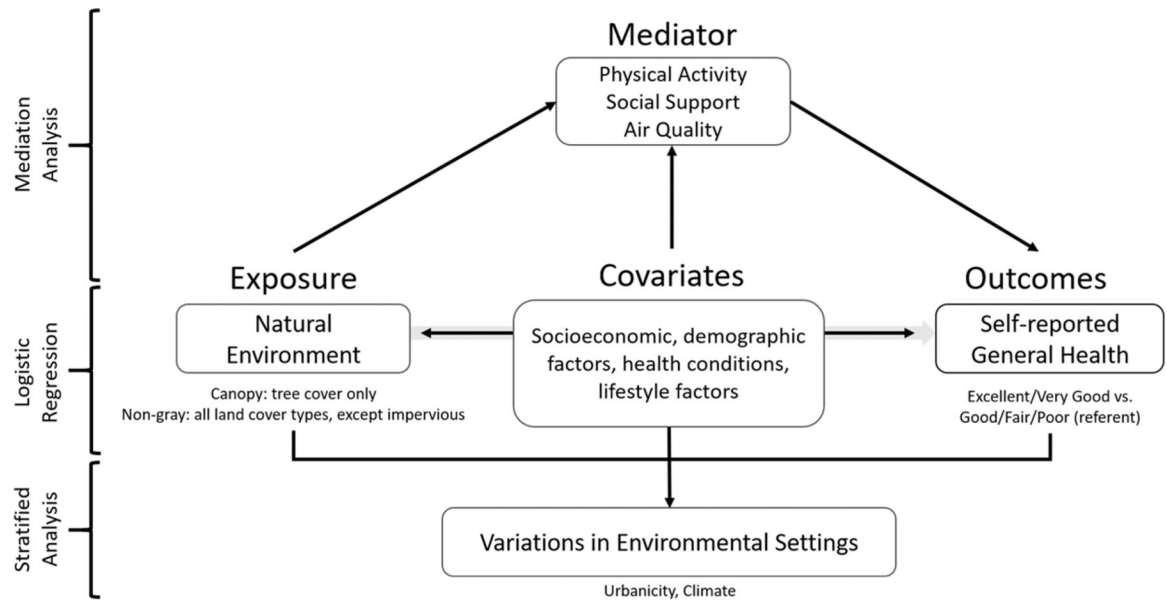
### Highlights

- Women living in a neighborhood with more natural features reported better SRGH.
- Stronger associations observed in urban areas and in continental climate.
- Social support and physical activity were stronger mediators than air quality.
- Observed effect is small but of public health importance.



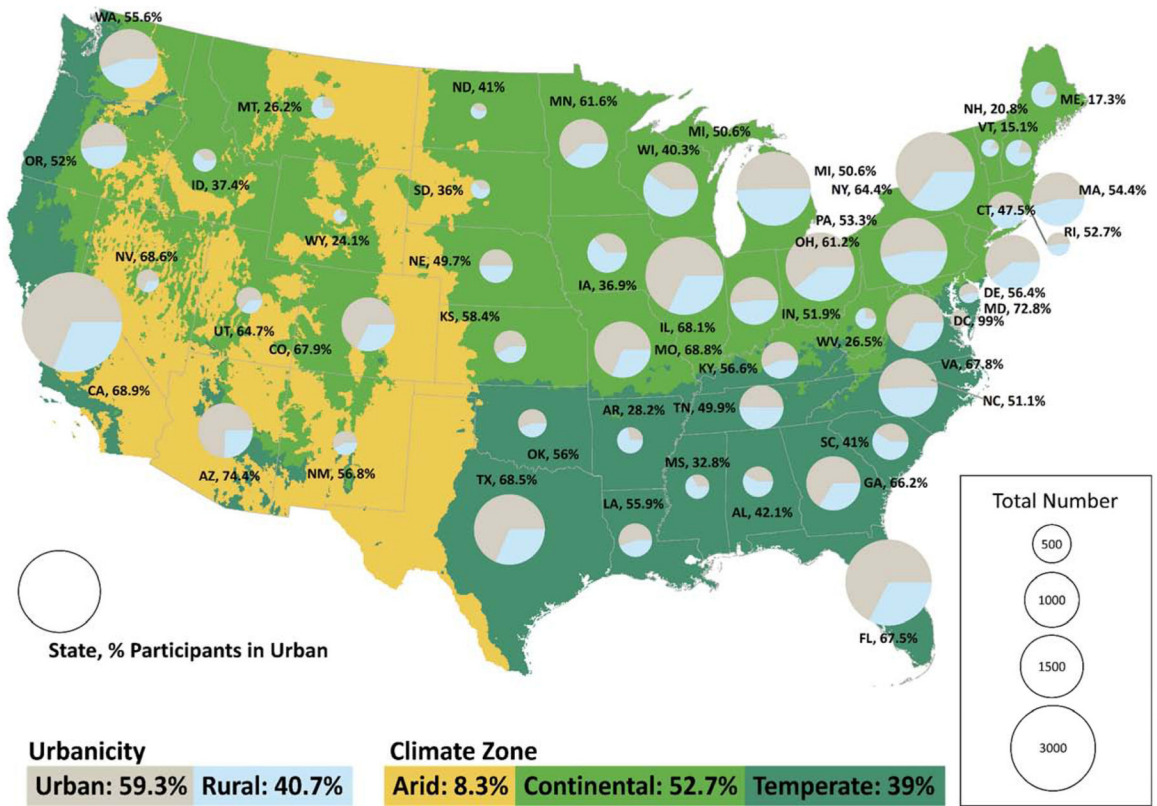
**Figure 1.**

a) Distribution of Sister Study participants in the conterminous United States. b) An example of participants with less than 95% coverage of their neighborhood analysis units.



**Figure 2.** Analytical approach for examining the associations between self-reported general health and natural environment.





**Figure 3.**  
Strata of urbanicity and climate zone.



**Table 1.**

Characteristics of participants by Self-reported general health.

<b>Health outcome - General Health</b>	<b>Good/Fair/Poor</b>	<b>Excellent/Very Good</b>	<b>N<sup>†</sup></b>
	10418(31.0%)	23188(69.0%)	33606
<b>Measures of Natural Environment</b>	<b>Mean (sd)</b>	<b>Mean (sd)</b>	<b>Mean (sd)</b>
Canopy at 250m	21.8 (20.3)	23.9 (21)	23.2 (20.8)
Canopy at 1250m	23.5 (19.3)	24.8 (19.5)	24.4 (19.4)
Non-gray at 250m	71.1 (19.7)	73.1 (19.1)	72.5 (19.3)
Non-gray at 1250m	74.4 (18.2)	76.0 (17.5)	75.5 (17.8)
<b>Covariates</b>	<b>Category/Continuous</b>	<b>N (%) / Mean (sd)</b>	<b>N (%) / Mean (sd)</b>
<b>Socioeconomic and demographic factors</b>			
Age	(Continuous)	58.9 (8.9)	57.8 (8.7)
Income	<\$20k	745 (7.2%)	511 (2.2%)
	\$20–50k	2859 (27.4%)	4091 (17.6%)
	\$50–100k	4139 (39.7%)	8937 (38.5%)
	\$100–200k	2184 (21.0%)	7313 (31.5%)
	\$200k+	491 (4.7%)	2336 (10.1%)
Job	Unable to Work	548 (5.3%)	198 (0.9%)
	Working	6397 (61.4%)	16300 (70.3%)
	Unpaid/Other Occupation	3473 (33.3%)	6690 (28.9%)
Education	High School or less	1897 (18.2%)	2909 (12.5%)
	Associate/College or so	3844 (36.9%)	7028 (30.3%)
	Bachelor	2510 (24.1%)	6818 (29.4%)
	Graduate	2167 (20.8%)	6433 (27.7%)
Race	Non-Hispanic White	8468 (81.3%)	20713 (89.3%)
	Non-Hispanic Black	1212 (11.6%)	1450 (6.3%)
	Hispanic	423 (4.1%)	524 (2.3%)
	Other	315 (3.0%)	501 (2.2%)
Area Deprivation Index (ADI)	(Continuous)	40.4 (25.6)	33.1 (23.3)
<b>Lifestyle</b>			35.3 (24.4)

Smoking	Never	5524 (53.0%)	13507 (58.2%)	19031 (56.6%)
	Former	3830 (36.8%)	8269 (35.7%)	12099 (36.0%)
	Current	1064 (10.2%)	1412 (6.1%)	2476 (7.4%)
Drinking	Never/Former	2500 (24.0%)	3400 (14.7%)	5900 (17.6%)
	Current Social	950 (9.1%)	1656 (7.1%)	2606 (7.8%)
	Current Regular	6968 (66.9%)	18132 (78.2%)	25100 (74.7%)
<b>Health Condition</b>				
Body Mass Index	(continuous)	29.8 (7.0)	26.5 (5.3)	27.6 (6.1)
Use of Depression Medication	No	7147 (68.6%)	19526 (84.2%)	26673 (79.4%)
	Yes	3271 (31.4%)	3662 (15.8%)	6933 (20.6%)
Use of Anxiety Medication	No	8271 (79.4%)	20677 (89.2%)	28948 (86.1%)
	Yes	2147 (20.6%)	2511 (10.8%)	4658 (13.9%)
<b>Survey Factor</b>				
Survey Season	Spring	539 (5.2%)	1110 (4.8%)	1649 (4.9%)
	Summer	3999 (38.5%)	9075 (39.1%)	13074 (38.9%)
	Fall	4625 (44.5%)	10790 (46.5%)	15415 (45.9%)
	Winter	1225 (11.8%)	2213 (9.5%)	3438 (10.2%)
<b>Strata</b>				
Urbanicity	Non-Urban	4388 (42.1%)	9279 (40.0%)	13667 (40.7%)
	Urban	6030 (57.9%)	13909 (60.0%)	19939 (59.3%)
Climate Zone	Arid climate	868 (8.3%)	1907 (8.2%)	2775 (8.3%)
	Continental climate	5346 (51.3%)	12378 (53.4%)	17724 (52.7%)
	Temperate climate	4204 (40.4%)	8903 (38.4%)	13107 (39.0%)
<b>Mediators</b>				
Walking	(continuous)	3.8 (5.3)	4.4 (5.5)	4.2 (5.4)
MVPA <sup>2</sup>	(continuous)	2.8 (3.8)	4.0 (4.1)	3.6 (4.1)
Social Support	(continuous)	13 (3.2)	14.1 (2.5)	13.7 (2.8)
Support Network	< 6 people	6266 (61.5%)	11974 (52.5%)	18240 (55.3%)
	>= 6 people	3921 (38.5%)	10813 (47.5%)	14734 (44.7%)
PM2.5	(continuous)	9.0 (1.8)	8.9 (1.8)	8.9 (1.8)
NO2	(continuous)	8.1 (4.4)	8.0 (4.3)	8.1 (4.3)

Statistics are means and standard deviation (sd) for continuous variables and number and percent (%) for categorical variables.

<sup>1</sup>Total number excluding missing values of General Health, covariates, and stratification factors (a total of 7,521 cases were missing).

Moderate-to-vigorous physical activity.

EPA Author Manuscript

EPA Author Manuscript

EPA Author Manuscript

**Table 2.**

Adjusted odds ratios (AOR) for reporting “Excellent/Very Good” versus “Good/Fair/Poor” general health for a 10% increase in natural environment.

	<b>AOR</b>	<b>Lower CI</b>	<b>Upper CI</b>	<b>C stat<sup>f</sup></b>
Canopy at 250m	<i>1.02</i>	<i>1.00</i>	<i>1.03</i>	<i>0.72</i>
Canopy at 1250m	<i>1.02</i>	<i>1.00</i>	<i>1.03</i>	<i>0.72</i>
Non-gray at 250m	<i>1.03</i>	<i>1.01</i>	<i>1.04</i>	<i>0.72</i>
Non-gray at 1250m	<i>1.03</i>	<i>1.01</i>	<i>1.04</i>	<i>0.72</i>

*Italic indicates that significance level is at 0.05.*

Model controlled for age, household income, employment status, educational attainment, race/ethnicity, smoking status, drinking status, body mass index, use of depression medication, use of anxiety medication, survey season, and area deprivation index

<sup>f</sup>Concordance statistic--a value above 0.7 indicates a good model.

**Table 3.**

Adjusted odds ratios (AOR) for reporting “Excellent/Very Good” versus “Good/Fair/Poor” general health given a 10% increase in natural environment, stratified by urbanicity, climate zone, and a combined factor of urbanicity and climate zone.

Stratified factor	Group	N <sup>1</sup>	Canopy at 250m	Canopy at 1250m	Non-gray at 250m	Non-gray at 1250m
Urbanicity	Non-Urban	13,677				
	AOR [95% CI] (mean, sd) <sup>2</sup>		1.00 [1.00, 1.03] (25.72, sd = 22.45)	1.00 [1.00, 1.03] (28.16, sd = 21.38)	1.02 [1.00, 1.04] (82.03, sd = 16.94)	1.04 [1.01, 1.07] (86.21, sd = 13.83)
Climate Zone	Urban	19,957				
	AOR [95% CI] (mean, sd)		1.03 [1.01, 1.05] (21.15, 19.16)	1.02 [1.00, 1.04] (21.6, 17.33)	1.03 [1.01, 1.05] (65.48, sd = 18.02)	1.02 [1.00, 1.05] (67.86, sd = 16.34)
Climate Zone	Arid	2,800				
	AOR [95% CI] (mean, sd)		0.93 [0.75, 1.16] (1.94, 4.08)	1.00 [0.74, 1.34] (1.69, 3.1)	1.00 [1.00, 1.05] (61.01, 18.24)	1.01 [1.00, 1.06] (65.07, 18.07)
Climate Zone	Continental	17,800				
	AOR [95% CI] (mean, sd)		1.04 [1.02, 1.06] (22.67, 18.73)	1.03 [1.01, 1.06] (24.2, 17.53)	1.04 [1.02, 1.06] (72.83, 19.69)	1.03 [1.01, 1.06] (76.07, 18.05)
Climate Zone	Temperate	13,182				
	AOR [95% CI] (mean, sd)		1.01 [1.00, 1.03] (27.89, 22.28)	1.01 [1.00, 1.03] (29.09, 20.17)	1.01 [1.00, 1.03] (73.55, 18.47)	1.02 [1.00, 1.04] (76.28, 16.8)
Urbanicity in Climate Zone	Non-Urban in Arid	863				
	AOR [95% CI] (mean, sd)		0.82 [0.57, 1.20] (2.06, 4.18)	0.88 [0.52, 1.48] (1.73, 3.16)	1.03 [1.00, 1.12] (73.53, 19.07)	1.05 [0.95, 1.15] (79.41, 16.69)
Urbanicity in Climate Zone	Continental	7,756				
	AOR [95% CI] (mean, sd)		1.02 [1.00, 1.05] (24.89, 21.2)	1.02 [1.00, 1.05] (27.61, 20.14)	1.04 [1.01, 1.08] (82.5, 16.89)	1.06 [1.02, 1.11] (86.89, 13.65)
Urbanicity in Climate Zone	Temperate	5,151				
	AOR [95% CI] (mean, sd)		1.00 [1.00, 1.01] (30.75, 23.32)	1.00 [1.00, 1.02] (33.22, 21.55)	1.00 [1.00, 1.02] (82.43, 16.55)	1.00 [1.00, 1.05] (86.06, 13.62)
Urbanicity in Climate Zone	Urban in Arid	1,973				
	AOR [95% CI] (mean, sd)		1.00 [0.75, 1.30] (1.89, 4.03)	1.05 [0.73, 1.51] (1.67, 3.08)	1.00 [0.93, 1.08] (55.57, 14.89)	1.02 [0.94, 1.10] (58.84, 14.82)

<b>Stratified factor</b>	<b>Group</b>	<b>N<sup>1</sup></b>	<b>Canopy at 250m</b>	<b>Canopy at 1250m</b>	<b>Non-gray at 250m</b>	<b>Non-gray at 1250m</b>
	<b>in Continental</b>	10, 044				
	AOR [95% CI]		<i>1.05 [1.02, 1.08]</i>	<i>1.05 [1.02, 1.09]</i>	<i>1.04 [1.02, 1.07]</i>	<i>1.03 [1.00, 1.06]</i>
	(mean, sd)		(20.99, 16.4)	(21.62, 14.74)	(65.49, 18.45)	(67.86, 16.59)
	<b>in Temperate</b>	8, 031				
	AOR [95% CI]		1.02 [1.00, 1.05]	1.01 [0.98, 1.04]	1.03 [1.00, 1.06]	1.02 [1.00, 1.05]
	(mean, sd)		(26.07, 21.39)	(26.45, 18.77)	(67.89, 17.36)	(70.06, 15.63)

*Italic indicates that significance level is at 0.05.*

<sup>1</sup>Total number in the stratum.

<sup>2</sup>Numbers refer to the mean value and standard deviation of natural environment in the stratum.

Model controlled for age, area deprivation index, household income, employment status, educational attainment, race/ethnicity, smoking status, drinking status, body mass index, use of depression medication, use of anxiety medication, and survey season.

Mediation effects by behavioral factors (walking and MVPA<sup>1</sup>), social factors (social support and support network), and air quality factors (PM<sub>2.5</sub> and NO<sub>2</sub>).

**Table 4.**

Mediator	Canopy at 250m		Canopy at 1250m		Non-gray at 250m		Non-gray at 1250m	
	%Mediated [95% CI]	%Mediated [95% CI]	%Mediated [95% CI]	%Mediated [95% CI]	%Mediated [95% CI]	%Mediated [95% CI]	%Mediated [95% CI]	
Walking	-0.97 [-3.29, 1.34]	-0.15 [-3.04, 2.74]	-1.54 [-3.42, 0.34]	-0.93 [-2.95, 1.09]				
MVPA	<i>10.06 [0.11, 20.01]</i>	13.52 [-0.89, 27.92]	<i>19.61 [7.93, 31.29]</i>	<i>23.58 [6.85, 40.3]</i>				
Social Support	9.49 [-1.05, 20.03]	10.75 [-1.84, 23.35]	<i>15.35 [5.22, 25.47]</i>	<i>15.09 [4.61, 25.56]</i>				
Support Network	6.80 [-1.18, 14.79]	3.73 [-2.80, 10.25]	3.59 [-0.03, 7.21]	1.30 [-2.11, 4.72]				
PM <sub>2.5</sub>	40.23 [-16.00, 96.45]	57.01 [-26.76, 140.78]	29.47 [-31.94, 90.89]	57.53 [-21.04, 136.11]				
NO <sub>2</sub>	5.57 [-2.84, 13.99]	6.68 [-3.87, 17.23]	6.15 [-12.16, 24.46]	8.00 [-17.31, 33.31]				

Italic indicates that significance level is at 0.05.

<sup>1</sup>Moderate-to-vigorous physical activity.