Research Article

Alexis Wilderman, Marcus Lam, Zhi-Yong Yin*

A pilot study exploring the relationship between urban greenspace accessibility and mental health prevalence in the City of San Diego in the context of socioeconomic and demographic factors

https://doi.org/10.1515/openhe-2021-0004
received August 28, 2021; accepted December 8, 2021

Abstract: The connection between urban greenspace and mental health is a robust but unsettled area of research in the public health and urban planning literatures. Inconsistent findings in prior studies are mostly due to differences in greenspace measurements and interrelations with socio-demographic factors. This study examines the relationships of mental health prevalence (MHP) with health prevention, socioeconomic and race-ethnicity factors, and proximity to greenspace at the census-tract level in the City of San Diego, California, using data from the CDC 500 Cities Project and US Census Bureau. We considered three greenspace proximity measures: distances to specified vegetation types, parks, and tree cover. Spearman's rank correlation showed that MHP was significantly correlated to distances to greenspace (rho = 0.480), parks (rho = 0.234), and tree cover (rho = 0.342), and greenspace proximity plus crime occurrence explained 37.8% of the variance in MHP in regression analysis. Further analysis revealed that socioeconomic status, race-ethnicity, and health prevention explained more than 93% of the variance in MHP, while greenspace proximity did not enter the regression model with statistical significance. We discovered that certain socioeconomic and race-ethnicity variables, such as proportion of Hispanic population, poverty, and regular checkup, may fully represent the effects of greenspace on MHP in the City of San Diego. Regression analysis for three subsections of the city suggested that different predictors of MHP should be considered in formulating intervention measures. Our results indicate the need to improve mental health conditions through a range of interventions that address the disparities experienced by racial-ethnic minorities and those in lower-socioeconomic classes.

Keywords: mental health prevalence, proximity to greenspace and parks, socioeconomic status, racial-ethnic minorities, census tracts

1 Introduction

There has been an increasing interest in understanding how the environment and human health are interconnected. Mental health is fundamental to human health and includes one's emotional, psychological, and social well-being [1]. According to the Centers for Disease Control and Prevention (CDC), mental illnesses are some of the most common health conditions in the United States (US) [1]. It has been estimated that more than 50% of people will be diagnosed with a mental illness in their lifetime [1]. Mental health and mental illness are points on a continuum with poor mental health lying somewhere in the middle [2]. There are many factors that can contribute to mental illnesses, including genetics, early adverse life experiences, and experiences related to chronic medical conditions [1]. Studies have suggested significant links between mental health and the built environment, including mental illness symptoms such as depression and anxiety and urban greenspace, but with mixed results [3–9]. This study examines mental health condition prevalence in relation to greenspace accessibility and socioeconomic and demographic factors, such as race-ethnicity.

***Corresponding author: Zhi-Yong Yin,** Environmental & Ocean Sciences, University of San Diego, San Diego, CA 92110, USA,
e-mail: zyin@sandiego.edu
Alexis Wilderman, Environmental & Ocean Sciences, University of San Diego, San Diego, CA 92110, USA,
e-mail: awilderman@sandiego.edu
Marcus Lam, School of Leadership and Education Sciences, University of San Diego, San Diego, CA 92110, USA,
e-mail: mlam@sandiego.edu

The study site is the City of San Diego, with a population of 1.4 million [10], with a unit of analysis at the census-tract level. Through our study, we hope to contribute to a deeper understanding of the extent to which the spatial distribution pattern of mental health conditions is influenced by accessibility to greenspace, socioeconomic factors, and race-ethnicity factors.

Urban greenspace broadly refers to open space with natural vegetation, including parks, gardens, yards, and urban forests [11], which can provide a place for social interactions and recreational activities to occur [3]. There is evidence that a positive relationship between greenspace and mental health exists, with evidence showing that greenspace can encourage social interaction, exercise, and provide a calming environment [12]. However, while the idea of greenspace and its associated positive benefits has been embraced by policymakers, organizations, and residents [4], the empirical evidence has been mixed [3–9]. Prior studies have presented conflicting results or have found no association between specific outcomes such as mental health and greenspace [3,5]. Studies in different geographical locations have found different associations between greenspace and a variety of health outcomes, suggesting that the relationship is dependent on environmental, cultural, and social contexts [3,6]. Additionally, studies have suggested that failing to acknowledge various neighborhood factors can potentially mislead conclusions concerning the greenspace-health relationship [7–9].

The rest of the paper examines a brief history of studies that have examined the relationship between urban greenspace and health, along with other factors that have been linked to mental health. The 'Data and methods' section presents the study area and the data sources analyzed, while the findings are presented in the 'Results' section. The 'Discussion' contains further examination of the findings with an in-depth examination of the limitations of this study.

1.1 A brief examination of the study site

We selected the City of San Diego as a case study for a variety of reasons. First, it is the second largest city in California and the eighth largest in the US with a rapidly growing and diverse population [10,13]. Additionally, due to its proximity to the busiest international border in the US, many residents of San Diego are "bi-national" in the sense that labor, commerce, and social activities criss-cross daily across the border [10]. Additionally, the County of San Diego has had the highest number of refugee arrivals in California for the past five years contributing to a more diverse but potentially underserved population [14]. The foreign-born population in San Diego, including both immigrants and refugees, represents 27.1% of the city's total population [15]. Finally, the unique geography of the City of San Diego, with its proximity to beaches and coastline, along with canyons, hiking trails, and other vegetation may serve to moderate the socioeconomic challenges facing the city and thus serves as an ideal case for this study.

1.2 The relationship between greenspace and mental health

Previous studies on the relationship between greenspace and mental health find mixed empirical results. One explanation for this variation in findings is that previous studies examined different measurements of greenspace, which were investigated at various spatial scales. For example, many studies examined the "greenness", or relative density and condition of vegetation in an area, which is commonly measured using the normalized difference vegetation index (NDVI) [5,7,9]. Measuring greenspace through greenness based on remote sensing data is a very common method [5]. A study conducted for the 496 major cities in the US found that cities with higher greenness had lower rates of poor mental health prevalence (MHP) [7]. Further, a study based in Wisconsin in the US found that a 25% higher NDVI was associated with a decrease in depression ($p < 0.01$) and anxiety ($p < 0.05$) [16].

Tree canopy cover or tree cover density has also been examined extensively in previous studies. Tree canopy cover data has been obtained from many various sources, and to our knowledge, there is not one specific database that was used consistently [7,16,17]. Overall, tree canopy cover was used to represent specific vegetation types such as trees or other forested environments — besides what is measured through NDVI or other greenness measures — and how they might be impactful on health [8,16]. For example, a study conducted in Australia found that increased tree canopy cover was associated with lower rates of mental distress [17]. On the other hand, Browning and Rigolon [7] used tree cover, in addition to NDVI, but found no significant relationship between tree canopy cover and mental health conditions, contradicting to their conclusion based on NDVI, and contradicting also to the results from the study mentioned above [16].

Another measure of greenspace is the proximity to greenspace [5]. Dzhambov et al. [9] found that distance to or access to the nearest greenspace was the third most common greenspace measure, examined in 19 (17.9%)

out of the 106 studies reviewed. Nutsford et al. [3] found that decreased distance to usable greenspace was associated with decreased anxiety and mood disorders. In contrast, other studies have used Euclidean distance to the nearest urban greenspace, including park, allotment, or recreational ground, but have found no direct association to mental health [8]. A cross-sectional study in Bulgaria found that there was no direct association between mental health and exposure to urban greenspace represented by all three greenspace measurements previously discussed, including NDVI, tree cover density, and Euclidean distance to urban greenspace [8].

1.3 Other factors of MHP

Many factors influence an individual's or community's mental health conditions and overall health outcomes due to dramatic health inequities in the US [7]. Socioeconomic and demographic factors, such as income, employment, gender, education, and race-ethnicity are well studied factors that influence mental health [18]. For example, it has been found that communities of color have less access to mental health services and receive lower quality care [2]. Although there are similar rates of mental disorders between communities of color and White communities, due to the lower utilization and poorer care that communities of color receive, the Office of the Surgeon General has estimated that communities of color have higher mental health needs [2].

When socioeconomic and demographic factors were included in studies examining the greenspace-health relationship, the effects of greenspace accessibility on health may be obscured or modified. For example, Lanza et al. [19] examined the census tracts of 25 US cities based on race, ethnicity, and income to understand how demographics moderated the relationship between vegetation and physical activity. It was found that there were mixed results between tree canopy cover and physical activity based on the census tract types, with Hispanic and low-income census tracts exhibiting increases in physical activity when tree canopy cover increased, but Black and low-income tracts and White and high-income tracts exhibiting decreases in physical activity when tree canopy cover increased. In contrast, there was a negative association between non-tree vegetation and physical activity in which a 10% increase in non-tree vegetation resulted in a decrease in physical activity for all census tracts regardless of demographics [19]. Additionally, Browning and Rigolon [7] found that race and ethnicity moderated the greenspace-health link. For example, it was found that in US cities with a greater percentage of non-Hispanic White residents, higher tree canopy cover was correlated with lower obesity rates; however, in US cities with smaller percentages of non-Hispanic White residents, higher tree canopy cover was correlated with higher obesity rates [7].

Despite the growing evidence of the positive impact of greenspace accessibility on improving mental health conditions, limited research exists on how the relationship is affected by socioeconomic factors and race-ethnicity. Therefore, the purpose of this study is to examine the greenspace-mental health relationship to achieve a better understanding to the associations between mental health, environment, socioeconomic status (SES), and demographics. There is also the need to further evaluate different measures of greenspace accessibility in such investigations. As such, we hypothesize that better accessibility to urban greenspace improves mental health conditions, but this relationship can also be influenced by various socioeconomic characteristics and race-ethnicity factors. Findings from this study or similar studies can help policymakers understand how accessibility to greenspace can influence the variation patterns of mental health conditions in a city, with considerations of socioeconomic and race-ethnicity disparities.

2 Data and methods

2.1 Study area

The City of San Diego is located in southern California, US (Figure 1) with 286 census tracts and 1,400,982 residents based on the American Community Survey (ACS) 2015 5-year average [20]. The city is situated in a semi-arid Mediterranean climate with warm and dry summers and cool and wet winters. The main vegetation types in the region include coastal sage scrub, chaparral, riparian forests, and grasslands [21].

2.2 Sources of data

2.2.1 MHP

We obtained census tract-level health data from the 2019 release of the CDC 500 Cities Project [22]. The 2019 data release is based on 2017 or 2016 model-based small area estimates (SAE), generated from the 2017 or 2016 Behavioral Risk Factor Surveillance System (BRFSS) data, 2010 census population data, and 2013–2017 or 2012–2016

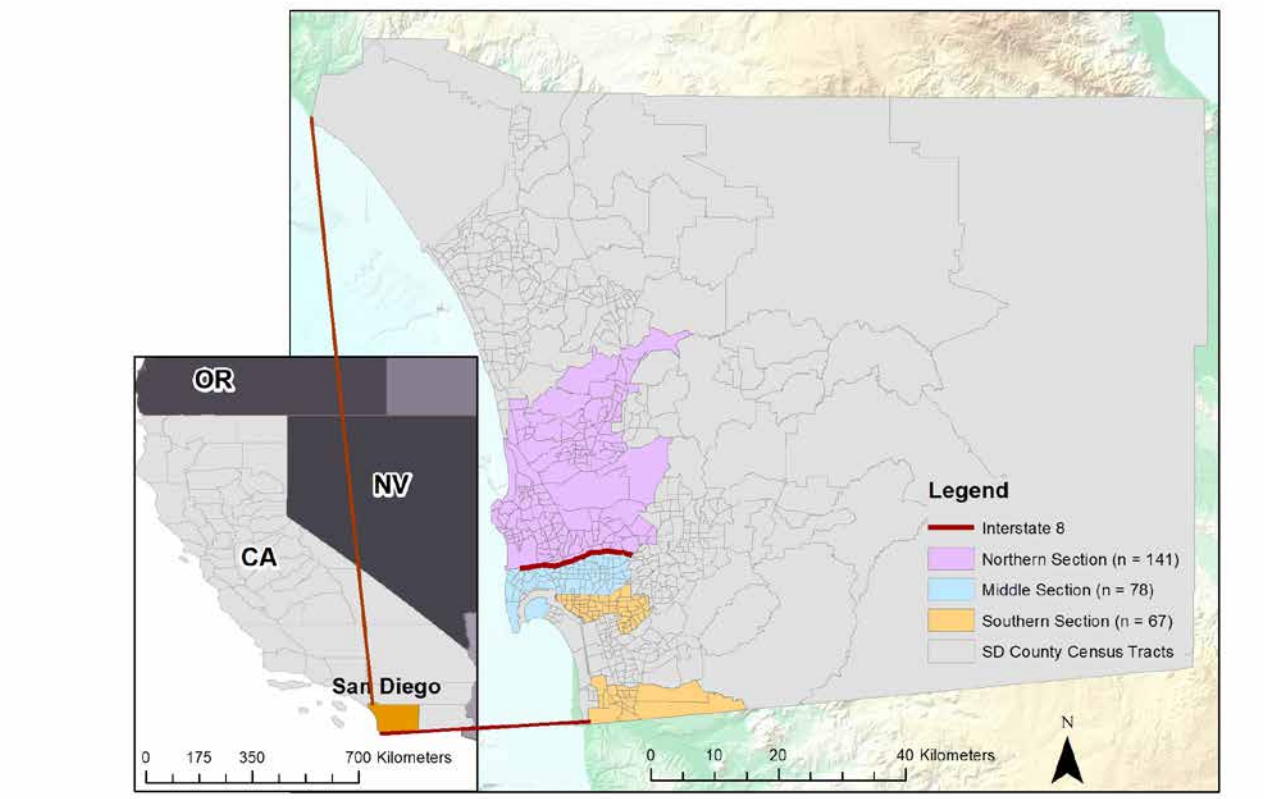

Figure 1: The City of San Diego's census tracts (n = 286) divided into three subsections: North, Middle, and South highlighted in light purple, blue and tan, respectively, within the County of San Diego in California, USA.

ACS estimates at the census tract level [22]. The CDC SAE contain data values that are indirectly estimated for a geographic area when direct survey estimates cannot produce valid and reliable estimates due to factors such as low response rate, limited sample sizes, or a lack of samples [23]. The SAE data can be used to indicate local health behavior or disease prevalence and it has been used in "secondary analyses to identify potential community-level factors that may be associated with health behavior or disease prevalence" [23]. Further, according to Kong and Zhang [23], the SAE data are more precise than those provided by direct survey-estimates at the census-tract level. More details about this dataset can be found in the Supplementary Part 1.

The MHP is measured as the percentage of respondents aged ≥18 years who reported that their mental health was not good in 14 or more of the past 30 days (Figure 2) [22]. Specifically, poor mental health conditions included reporting symptoms of stress, depression, and problems with emotions for 14 or more of the past 30 days [22]. The census tracts could represent either neighborhoods or local communities with distinct socioeconomic and racial-ethnic characteristics and offer a higher spatial resolution than city- or county-level data. Census tract boundaries are consistent with prior studies in the public health literature that explored the spatial relationships between neighborhood characteristics and various health outcomes, such as child well-being [24], healthy food access [25], and measures of general health [26].

2.2.2 Urban greenspace accessibility

In many previous studies, especially those based on city- or county-level data, greenspace accessibility was determined by proportions of areas of forest or average NDVI values summarized by the areal units [7,27]. At the census-tract level, however, such measures may underestimate the influence of greenspace located outside but still very close to a given census tract. In this current study, we decided to use proximity to represent greenspace accessibility: distances to greenspace as defined by vegetation types, established parks, and locations with significant amount of tree canopy coverage. Vegetation type data were obtained from San Diego's Regional GIS Data Warehouse Open Data Portal [28]. The greenspace was specifically defined as areas with any of the following vegetation types: Forest; Grasslands, Vernal Pools,

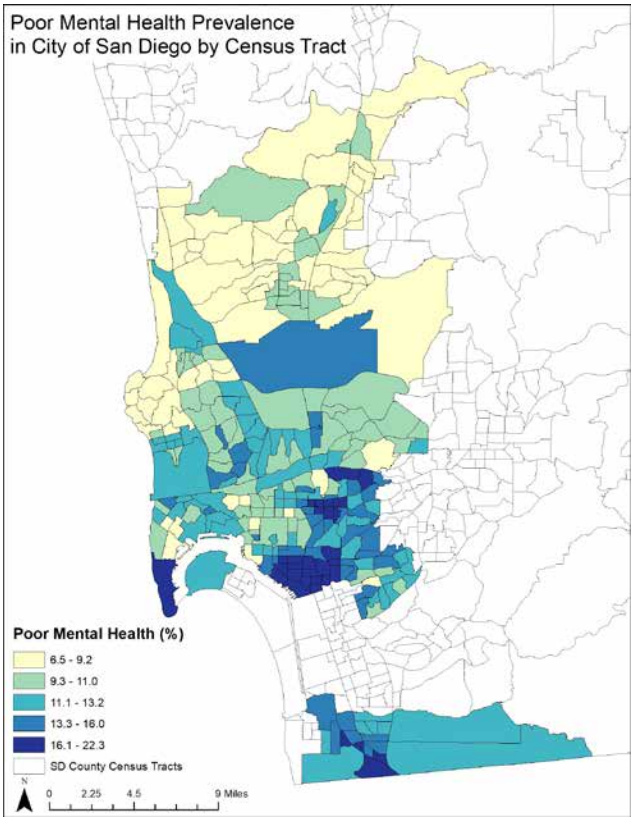

Figure 2: Mental health prevalence in the City of San Diego, with values displaying the percent of residents in a census tract that reported experiencing poor mental health conditions for 14 or more of the past 30 days. Data obtained from the CDC's 500 Cities Project [22].

Meadows, and other Herb Communities; Woodland; Riparian and Bottomland Habitat; Bog and Marsh; Dune Community; and Scrub and Chaparral. The GIS dataset of urban parks contains boundaries of public parks in San Diego and was obtained from the same source [28]. To better represent the quality of greenspace, we obtained the tree canopy cover data from the US USDA/Forest Service [29], based on Landsat imagery classification for pixels with different proportions of tree canopy cover. We focused on pixels with canopy cover proportions ≥30% since previous studies have shown that there are significant mental health benefits once tree cover reaches a minimum of 30% [17,30] and the highest percentage of tree canopy cover in the City of San Diego was 69% in this dataset. More details about the GIS data of greenspace and urban parks can be found in the Supplementary Part 1.

The proximity to urban greenspace was measured by calculating the mean distances (in feet) to the polygons of vegetation and parks, and pixels of tree canopy cover ≥30% using the Euclidean Distance tool and Zonal Statistics tool in ArcGIS Desktop 10.8 (ESRI, Redlands, CA, USA). As an example, Figure 3 shows the procedure of creating the distance surface to the vegetation polygons and the final summarized mean distance values by census tracts in the City of San Diego. English units were used in this process, but conversion factors are provided in the caption of Figure 3. These steps were repeated for the tree canopy cover and parks data to create distance surfaces that were summarized as mean distances by the census tracts (Supplementary Figure 1).

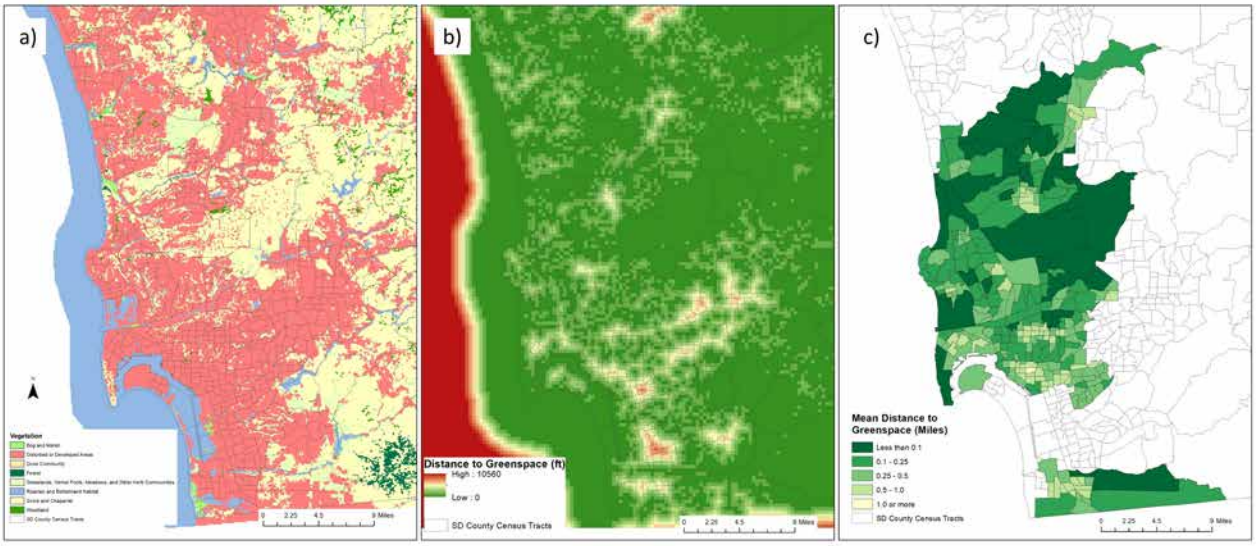

Figure 3: Creation of *distance to greenspace* in ArcGIS using (a) vegetation category data obtained from SANGIS.org [28], (b) Euclidean distance tool of ArcGIS to calculate distance to vegetation polygons, and (c) the Zonal Statistics tool to calculate mean distances for census tracts in the City of San Diego (conversion factors: 1 ft = 0.3048 m, 1 mile = 1.609 km).

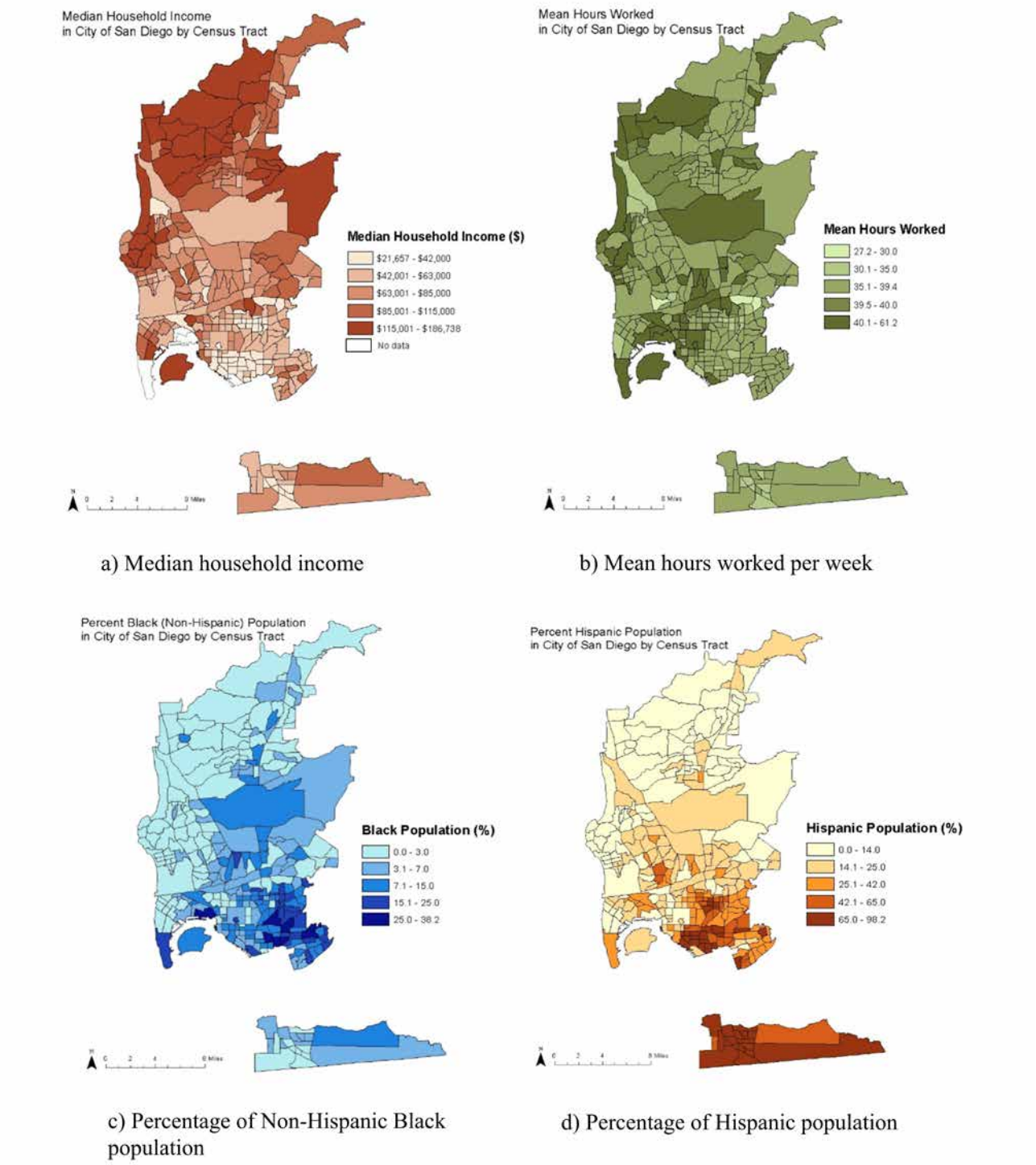

Figure 4: Examples of SES and race-ethnicity data obtained from the US Census Bureau's ACS database for the City of San Diego by census tracts.

2.2.3 Socioeconomic, race-ethnicity and crime density data

We considered a large variety of variables that may serve as stressors to cause hardship and poor mental health conditions. SES variables and demographic characteristics were collected from the US Census Bureau's ACS for 2011–2015 (5-year estimates) [7]. For SES variables, we included *median household income* (Figure 4a), *educational attainment* or *percent degree*, *housing value*, and *hours worked per week* (as a proxy of full-time employment) (Figure 4b). For demographic characteristics, we included race and

ethnicity data (Supplementary Part 1), including *percent non-Hispanic Black* (Figure 4c) and *Hispanic* population (Figure 4d). Previous studies have detailed the existing health inequalities between people of different races, ethnicities, gender, and social classes in the US, with evidence showing that racial, ethnic, and socioeconomic disparities in health widened since the 1980s [31]. Additionally, *percent uninsured* and *percent with routine checkups* were obtained from the CDC's 500 Cities Project. Percent uninsured is a variable closely related to SES factors as income and employment conditions often determine whether a household can afford health insurance coverage. On the other hand, percent of population with regular checkups was used as a proxy of mental health preventive measure.

Crime occurrence was also included in this study as a factor that may influence both MHP and certain SES variables, especially property value [32,33]. Previous studies have also shown that fear of crime can decrease a population's utilization of greenspace and parks [34]. Crime data was recorded over a 180-day period, which was obtained in October 2019 from the San Diego Association of Governments (SANDAG) Automated Regional Justice Information System (ARJIS) (Supplementary Part 1) [35]. *Crime density* is the number of crime occurrence per unit area (square mile) in each census tract and can be used to represent the relative spatial distribution of crimes in the City of San Diego. Crime density was calculated using the Kernel Density tool with a 1-mile (1.609 km) bandwidth and then summarized by the Zonal Statistics tool to get mean density values for census tracts in ArcGIS Desktop 10.8 (Figure 5).

We categorized these potential factors of MHP into different groups. According to previous studies [7,36], we decided that household income and educational attainment should both be considered as important SES factors. We also identified the percentage of uninsured population as an important SES factor, while the regular checkups served as a proxy of overall healthcare-seeking willingness or preventive measure (health prevention). The SES category also includes employment conditions (mean hours worked per week) and median housing value. Considering the shortage of affordable housing as an important stressor for the residents of San Diego [37], we calculated an index of housing affordability as the median housing value divided by the median household income, which is equivalent to the number of years to purchase a house of the median value with the median household income for a given census tract, and we included it in the SES category. The race-ethnicity category includes the percentages of the predominant minority ethnic groups: Hispanic and non-Hispanic Blacks. The list of the variables used in this study and their definitions are presented in Table 1.

2.3 Statistical analyses

Based on previous studies, e.g., [7,36], and examinations on how accessibility to greenspace, poor mental health, and other factors are related, a conceptual model was built including all potential factors of MHP considered in this study (Figure 6), serving as the basic construction of hypotheses. It displays the hypothesized direct and indi-

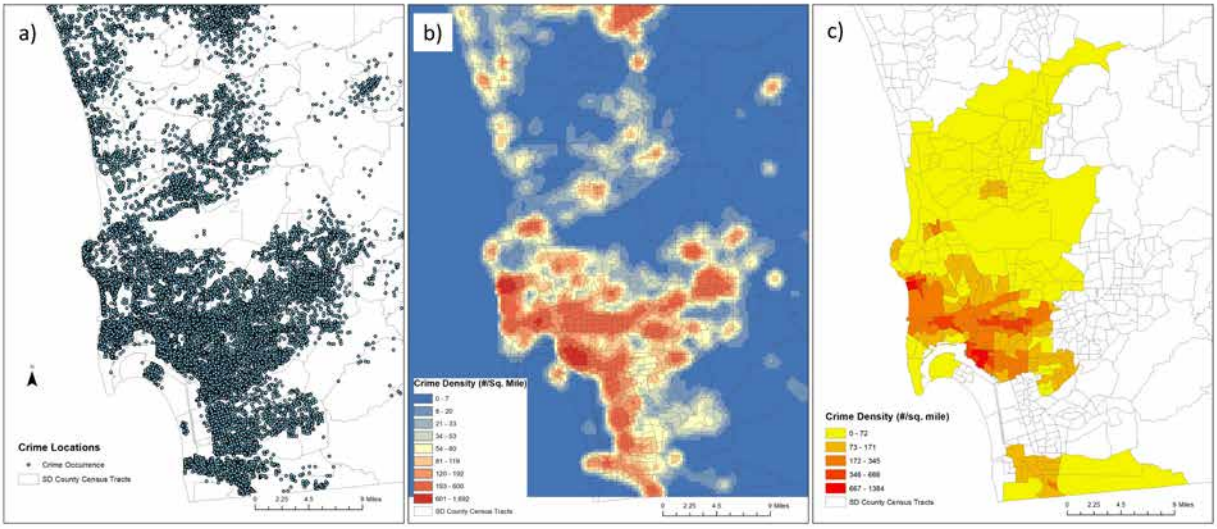

Figure 5: Creation of *crime density* in ArcGIS using (a) crime occurrence points in San Diego County from SANDAG [35], (b) kernel density tool to create density surface using 1-mile search radius for San Diego County, and (c) the Zonal Statistics tool to calculate mean crime density for census tracts in the City of San Diego (conversion factors: 1 mile = 1.609 km, 1 mile^2 = 2.589 km^2).

rect effects on poor mental health conditions by the aforementioned factors. Then Spearman's rank correlation was performed to test all hypotheses illustrated in the conceptual model and to determine the statistical associations between all variables and mental health conditions. Rank correlation was used to avoid the possible violation of the linear relationship assumption of the Pearson's product moment correlation [38].

Based on the results of correlation analysis, we performed regression analysis to examine the effects of multiple measures of greenspace accessibility on MHP. Next, following the approach in which the CDC SAE dataset was constructed [23], we first built a regression model that provides good estimates of MHP with the potential SES, health prevention, and race-ethnicity factors, carefully selected based on previous studies and the conceptual model (Figure 6). We used the stepwise least-square regression method to examine the sequence in which the independent variables entered the model, indicating their levels of relevance to MHP [39]. In this process, we ensured that all independent variables included in the final model are statistically significant at the 0.05 level with signs of the regression coefficients representing the same relations to MHP as in the rank correlation analysis. For the final model, we also used the standardized regression coefficients (Beta coefficients) to represent the relative contributions made by the independent variables. Because some of the SES and race-ethnicity variables may be highly interrelated, we monitored and managed the

Table 1: Descriptions of variables and data sources (variable names displayed in statistical analysis results).

Variable	Description	Source[1]
Mental health prevalence (MHP)	Percentage of residents who reported poor mental health 14 or more out of the last 30 days	CDC
Distance to greenspace ("greenspace")	Mean distance to greenspace as specified by polygons of different vegetation types	SANGIS/SANDAG
Distance to tree canopy cover ("tree cover")	Mean distance to tree canopy cover (30%–69% as tree canopies within a 30 m x 30 m pixel)	USDA/FS
Distance to parks ("parks")	Mean distance to polygons of parks	SANGIS/SANDAG
Median household income ("median income")	Census tract median household income in dollars	ACS
Percent degree ("degree" or "educational attainment")	Percentage of residents aged 25 and older with bachelor, associates, masters, professional, or doctoral degree	ACS
Median housing value ("housing value")	Median value of house in dollars	ACS
Mean hours worked per week ("hours worked")	Mean usual hours worked per week for residents in each census tract	ACS
Lack of affordable housing ("affordability")	Time required to purchase housing of median value by median household income in years	ACS
Percent below poverty ("poverty")	Percent of population below poverty line	ACS
Non-Hispanic Black ("Black")	Percentage of non-Hispanic Black residents	ACS
Hispanic ("Hispanic")	Percentage of Hispanic residents with any race	ACS
Density of crime occurrence ("crime density")	Number of crimes per square mile averaged per census tract	SANDAG
Percent without health insurance ("uninsured")	Percentage of residents who reported having no current health insurance coverage. Serves as a proxy of accessibility to healthcare	CDC
Percent with routine health checkups ("regular checkup")	Percentage of residents who reported seeing a doctor for a routine checkup in the last year. Serves as a proxy of health condition prevention	CDC

[1] CDC is Centers for Disease Control and Prevention. ACS is the American Community Survey run by the US Census Bureau. SANDAG is San Diego Association of Governments. SANGIS/SANDAG is the joint collaboration between San Diego Geographic Information Source and San Diego Association of Governments. USDA/FS is US Department of Agriculture/Forest Service.

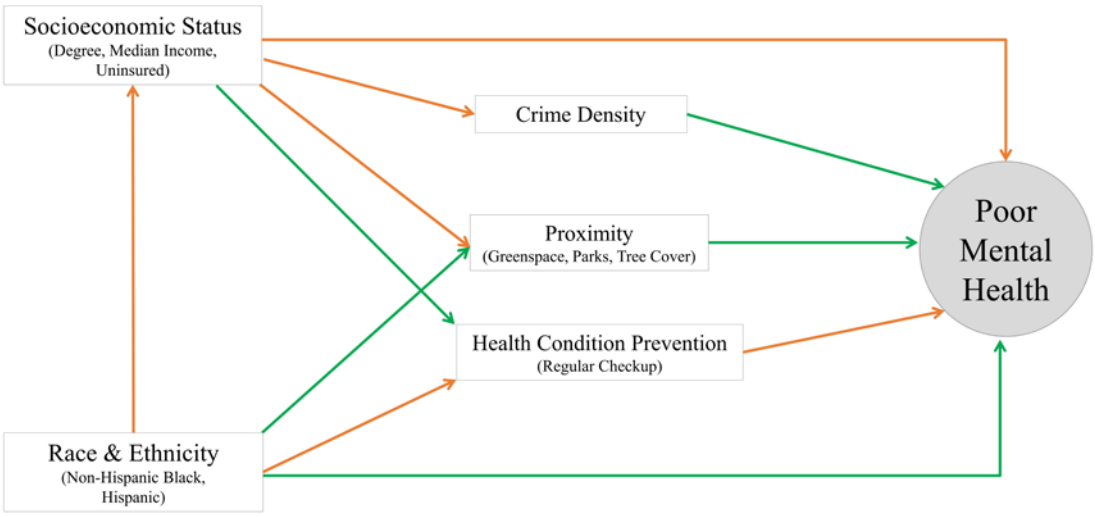

Figure 6: The conceptual model built to serve as the basic construction of hypotheses. Orange paths represent hypothesized negative correlations and green paths represent hypothesized positive correlations. Only representative variables are displayed.

collinearity among the independent variables measured by the variance inflation factor (VIF) [39].

Once the final model based on the SES and race-ethnicity variables was obtained, we would add the variables representing accessibility to urban greenspace/parks/tree cover and crime occurrence and examine how much contribution they can make to predict the variation in MHP with statistical significance. If they enter the model, it means that these factors have direct effects on MHP; otherwise, it is likely that their effects have been represented by other variables in the model already and, therefore, they only have indirect effects on MHP.

Since it has been found that greenspace's effects on mental health differ for areas with different SES and racial-ethnic compositions [7], we further examined how the model would perform in different geographical settings. Studies show that the City of San Diego continues to have high levels of racial and ethnic segregation even though the minority population has continued to increase. For example, from 1990 to 2000, the population of non-Hispanic White residents decreased in the southeastern portion of the city which is where minority populations increased the most [8]. Additionally, the City of San Diego has been ranked in the top ten US cities for the highest levels of overall educational segregation and highest levels of overall occupational segregation [40]. To consider the spatial patterns of the SES factors and racial-ethnic minority population distributions, we divided the City of San Diego into three different geographical sections: the most southern part of the City of San Diego, which includes those census tracts south of Downtown San Diego (n = 67), the middle part of the city between the Downtown and Interstate 8 or I-8 (n = 78), and the northern part of the city, which is everything north of I-8 (n = 141) (Figure 1) [41]. Generally speaking, the southern section is dominated by communities of minority ethnic groups such as Hispanics with relatively high poverty levels and low educational attainments, the middle section contains communities that are multi-ethnic with highly diverse SES conditions, while the northern section is dominated by non-Hispanic White population with the lowest poverty level among the three [41].

All statistical analyses were performed using IBM SPSS version 26, while all mapping and spatial analyses were performed using ArcGIS version 10.8 (ESRI, Redlands, CA). To avoid potential issues due to irregular values and missing data, we removed all census tracts with populations fewer than 200 for whom the ACS's poverty data were determined in all statistical analyses (6 tracts out of 286 were removed by this measure).

3 Results

3.1 Descriptive statistics

Our sample of census tracts (n = 280 after removal of tracts with <200 population) varied widely across all variables (Table 2). MHP representing percentage of *Poor Mental Health* ranged from 6.5% to 19.1%. The mean distances to greenspace, parks, and tree cover throughout

Table 2: Descriptive statistics of the variables used in analysis for census tracts in the City of San Diego (n = 280 after removal of tracts with fewer than 200 people unless otherwise specified).

	Minimum	Maximum	Mean	Std. deviation
Mental health prevalence (%)	6.5	19.1	11.6	2.9
Median household income ($)	$21,657	$186,738	$72,096	$32,922
Percent degree (%)	5.5	88.2	49.4	22.5
Hours worked (hrs/wk)	27.2	52.6	38.0	2.5
Housing value ($) (N = 275)	$124,500	$1,720,900	$482,256	$247,569
Affordability (years) (N = 275)	2.2	17.7	7.0	2.5
Percent below poverty (%)	1.0	52.6	15.3	11.9
Percent White (%)	0.8	97.9	46.6	27.7
Percent Black (%)	0.0	38.2	6.0	6.5
Percent Hispanic (%)	3.8	98.3	29.2	24.5
Regular checkup (%)	55.5	84.6	66.9	3.7
Percent uninsured (%)	4.0	32.5	11.8	7.2
Crime density (#/mi^2)	2.0	1384.3	169.9	213.9
Distance to greenspace (ft)	0.0	5352.7	1715.2	1173.1
Distance to tree cover (ft)	73.2	2759.9	558.4	412.2
Distance to parks (ft)	0.0	8048.5	1935.8	1265.7
Total population (# people)	1572	19414	4862	2120

the City of San Diego census tracts ranged from 0 miles to more than 1 mile (maximum of 5352.7 ft for greenspace, 8048.5 ft for parks, and 2759.9 ft for tree cover, 1 ft = 0.3048 m). *Median household income* ranged from $21,657 to $186,738. *Percent degree* representing educational attainment ranged from 5.5% to 88.2%. *Crime density* ranged from 2 crime incidents per square mile to more than 1,300 per square mile. The distribution of these variables often clustered geographically across the City of San Diego. For example, *MHP* has the pattern of high values in the south and low values in the north of the city (Figure 2), *crime density* was highest in the mid-western portion of the city (Figure 5), both *percent Hispanic* and *percent Black* were highest in the south-southeast of the city (Figure 4c and d), *median income* was highest in the north (Figure 4a), while *mean hours worked* did not seem to have a distinct pattern but with the highest in the western portion of the city (Figure 4b).

3.2 Correlation analysis and model structure

In order to test the hypotheses in the initial conceptual model (Figure 6), we first ran correlation analysis on the MHP and its potential factors using Spearman's rank correlation. The results indicate that all hypotheses in the conceptual model were confirmed with statistical significance, although the strengths of the apparent relationships varied greatly from 0.224 between the lack of housing affordability and MHP to 0.916 between the percentage of uninsured and MHP (Table 3). It can be seen that MHP is positively correlated with distances to greenspace (rho = 0.480), tree cover (rho = 0.342), and parks (rho = 0.234) with statistical significance at the 0.01 level, indicating that the longer the distances to greenspace and parks, the poorer the mental health conditions. However, MHP is much more strongly correlated with most SES variables (up to 0.916 for uninsured), preventive measure (rho = -0.594 with regular checkup), and race-ethnicity variables (rho = 0.842 and 0.578 with percentages of Hispanic and Black, respectively). There is also a moderately strong correlation with crime density (rho = 0.585). Additionally, it can be observed that many of these variables are intercorrelated to each other. For example, median household income is strongly correlated with percentage of uninsured, educational attainment (percent degree), median housing value, percent below poverty, and percentage of Hispanic population, and moderately correlated with accessibility to greenspace, health prevention, hours worked per week, crime occurrence, and percentage of Black population. Figure 7 is the updated conceptual model to illustrate the factors of MHP, showing the link-

ages between different groups of variables using the rank correlation results.

3.3 Regression modeling

To examine the combined effects of greenspace accessibility on mental health conditions in San Diego, we preform stepwise regression analysis using all greenspace proximity variables and the crime density, assuming that crime occurrence may negatively impact the usage of greenspace. Table 4a shows that the model explains 37.8% of the variance in MHP with all variables included in the model being statistically significant and representing the same relationships revealed by correlation analysis (Table 3; Figure 7). The distance to all parks did not enter the model, which also had a lower correlation coefficient than the other two greenspace proximity variables (Table 3).

Based on previous studies [7,23,36] and results of correlation analysis above, we assumed that median household income, educational attainment, lack of affordable housing, percent of uninsured population, poverty, and regular checkups (prevention) are direct factors of MHP, which either limit people's capability of seeking healthcare services or serve as stressors, while the race-ethnicity factor may have both direct and indirect effects on MHP as illustrated in the conceptual model (Figure 7). Therefore, we conceptualized the following general model structure:

MHP = f [uninsured, regular checkup, median income, degree, affordability, hours worked, poverty, race-ethnicity] (1)

Before we constructed the regression model of MHP using the aforementioned variables, we first examined the collinearity among the potential independent variables (Supplementary Part 2). Table 4b contains the final regression model for MHP after removing the independent variables with high VIF values (Supplementary Part 3). Overall, the model explained more than 93% of the variance in MHP (R^2 = 0.935; R_{adj}^2 = 0.933; F = 513.9; Sig. = <0.001), with all regression coefficients representing the same relationships with MHP as indicated by the earlier correlation analysis (Table 3). Figure 8 shows the predicted MHP values (after the antilog of the predicted lg_MHP values) plotted against the observed MHP values for the 257 tracts with valid estimates. It can be seen that the predicted values match well with the observed values in a strong linear relationship, with a regression coefficient (trend line slope) close to 1.0 and the intercept term close to zero. The scatter pattern does not display significant heteroscedasticity [42]. We also evaluated the residual to ensure that it is likely to be normally distributed (Supplementary Figure 2).

Using the same independent variables in Table 4b as the basis, we further added the variables of greenspace proximity (distances to greenspace, tree cover, and parks) and crime occurrence to see whether these factors have

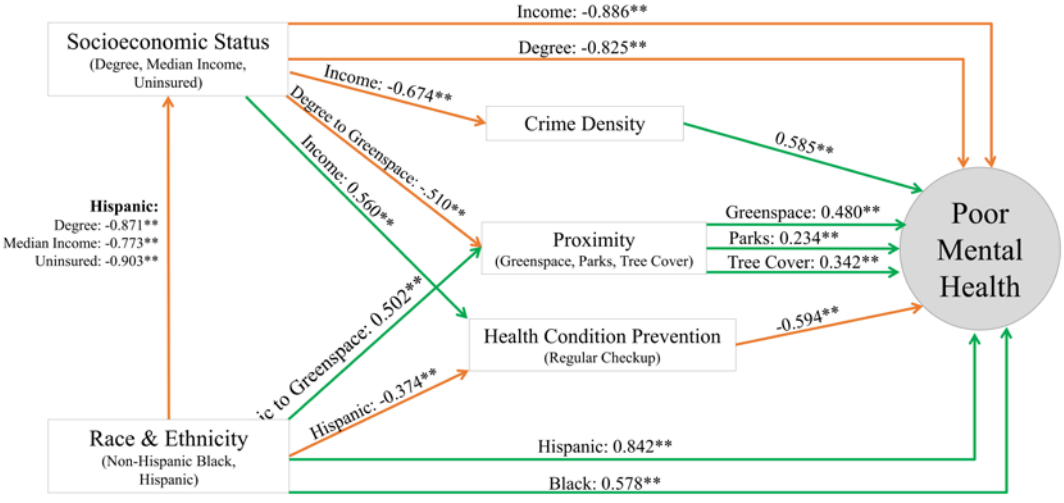

Figure 7: Examples of statistically significant linkages are shown as rank correlation coefficients on the conceptual model (Figure 6). Orange paths represent hypothesized negative correlations and green paths represent hypothesized positive correlations, in which all hypotheses from the conceptual model were confirmed. Only representative variables are displayed. * Statistically significant at the 0.05 level, ** significant at the 0.01 level.

Table 3: Spearman's rank correlation coefficients between the mental health prevalence (MHP) and its covariates. Sample size varied between 275 and 280 due to missing values and removal of tracts with population fewer than 200. * Statistically significant at the 0.05 level, ** significant at the 0.01 level.

Variables	MHP	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
MHP	1.000														
1. Greenspace	0.480**	1.000													
2. Tree cover	0.342**	0.419**	1.000												
3. Parks	0.234**	0.465**	0.245**	1.000											
4. Regular checkup	-0.594**	-0.212**	-0.170**	-0.154*	1.000										
5. Percent uninsured	0.916**	0.530**	0.397**	0.264**	-0.437**	1.000									
6. Percent degree	-0.825**	-0.510**	-0.371**	-0.236**	0.220**	-0.925**	1.000								
7. Median income	-0.886**	-0.558**	-0.282**	-0.280**	0.560**	-0.825**	0.739**	1.000							
8. Housing value	-0.747**	-0.531**	-0.343**	-0.264**	0.321**	-0.845**	0.787**	0.745**	1.000						
9. Affordability	0.224**	0.112	-0.083	0.073	-0.298**	0.014	0.016	-0.372**	0.287**	1.000					
10. Percent below poverty	0.823**	0.502**	0.257**	0.227**	-0.581**	0.743**	-0.587**	-0.844**	-0.631**	0.352**	1.000				
11. Hours worked	-0.569**	-0.416**	-0.274**	-0.218**	0.155**	-0.607**	0.580**	0.536**	0.558**	-0.035	-0.532**	1.000			
12. Crime density	0.585**	0.535**	0.219**	0.187**	-0.622**	0.470**	-0.339**	-0.674**	-0.388**	0.446**	0.680**	-0.221**	1.000		
13. Percent Hispanic	0.842**	0.502**	0.324**	0.229**	-0.374**	0.903**	-0.871**	-0.773**	-0.761**	0.067	0.657**	-0.536**	0.454**	1.000	
14. Percent Black	0.578**	0.341**	0.248**	0.200**	-0.159**	0.634**	-0.614**	-0.529**	-0.603**	-0.112	0.436**	-0.225**	0.310**	0.552**	1.000

Table 4: Results of stepwise regression using log-transformed MHP (lg_MHP) as the dependent variable (N = 257 after removal of census tracts with population fewer than 200 and missing values due to log-transformation). Column "B" is the regression coefficients and "Beta" is the standardized regression coefficients. Part a: Results using all greenspace proximity variables (distances to greenspace, tree cover, and parks) and crime density (log-transformed) as the independent variables. Part b: Results using the SES and race-ethnicity variables (greenspace proximity variables did not enter the final model).

Part a.	B	Std. error	Beta	t	Sig.	Tolerance	VIF
(Constant)	0.522	0.058		9.030	0.000		
lg_Crime density	0.091	0.012	0.429	7.273	0.000	0.653	1.530
lg_Tree cover	0.083	0.018	0.227	4.588	0.000	0.928	1.078
lg_Greenspace	0.042	0.018	0.140	2.334	0.020	0.628	1.593
Part b.	**B**	**Std. error**	**Beta**	**t**	**Sig.**	**Tolerance**	**VIF**
(Constant)	4.283	0.223		19.198	0.000		
lg_Hispanic	0.051	0.011	0.169	4.697	0.000	0.201	4.982
lg_Poverty	0.046	0.008	0.163	5.934	0.000	0.346	2.892
lg_Degree	-0.154	0.013	-0.404	-11.476	0.000	0.210	4.763
lg_Checkup	-1.241	0.098	-0.257	-12.692	0.000	0.633	1.579
lg_Hours worked	-0.572	0.071	-0.153	-8.051	0.000	0.717	1.394
lg_Black	0.024	0.004	0.121	6.464	0.000	0.742	1.348
lg_Affordability	0.074	0.013	0.101	5.546	0.000	0.778	1.286

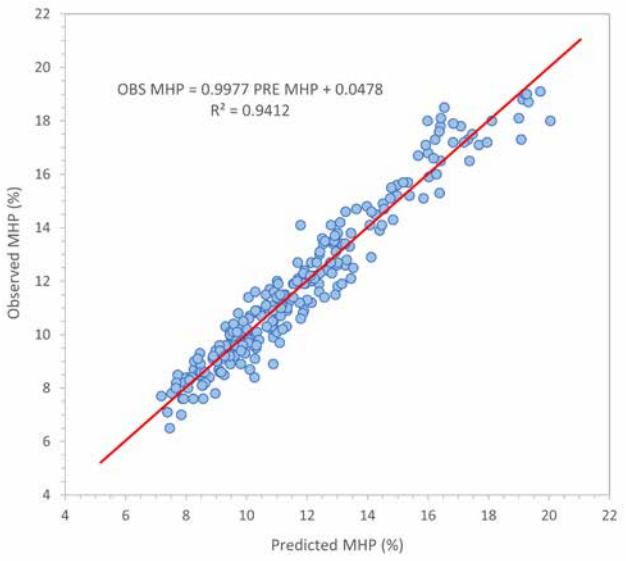

Figure 8: Observed and predicted mental health prevalence (MHP) values after antilog of the predicted values by the regression model presented in Table 4b.

any direct effects on MHP. However, none of these variables would enter and remain in the model with statistical significance at the 0.05 level. Since all of them are statistically correlated with MHP as shown in Table 3, we can conclude that the effects of accessibility to greenspace and crime occurrence on mental health conditions, if any, are most likely represented by these SES and race-ethnicity factors in the model.

To explain the results that proximities to greenspace and parks were significantly correlated with MHP (Table 3), but failed to enter the regression model, we performed partial correlation analysis between MHP and distances to greenspace, tree cover, and parks to test whether other socioeconomic variables can represent the effects of the proximities to greenspace and parks in explaining MHP. In this analysis, we examined the first order relationship between MHP and the three greenspace proximity variables using the SES and race-ethnicity variables as the control variable one at a time (Table 5). We discovered that a number of variables made the original correlations of distances to greenspace and parks to MHP (Table 3) statistically insignificant or even with reversed signs (i.e., from positive to negative, as shown in Table 3). We label these variables "Similar Variables" (top panel in Table 5) as they can replace the effects of accessibility to greenspace and parks on MHP in regression analysis, while those "Dissimilar Variables" (bottom panel in Table 5) do not have this effect. In other words, since the final regression model (Table 4b) contained several "Similar Variables", they prevented the greenspace proximity variables from entering the model with statistical significance to help further explain the variations in MHP. Using a subset of the Dissimilar Variables in combination with the greenspace proximity variables, we were able to build a regression

Table 5: Partial correlation coefficients between MHP and distances to greenspace (lg_greenspace, lg_tree cover) and parks (lg_parks) with different types of control variables (similar and dissimilar variables; see text for explanation).

Variable types	Control variables	lg_Greenspace	lg_Tree cover	lg_Parks
Similar variables	lg_Uninsured	-0.031	-0.096	0.014
	lg_Degree	0.112	0.025	0.087
	lg_Income	-0.114	0.217**	-0.052
	lg_Poverty	0.116	0.272**	0.088
	lg_Hispanic	0.077	0.103	0.083
Dissimilar variables	lg_Checkup	0.467**	0.263**	0.217**
	lg_Housing value	0.187**	0.164**	0.081
	lg_Affordability	0.466**	0.367**	0.196**
	lg_Hours worked	0.371**	0.312**	0.168**
	lg_Black	0.388**	0.219**	0.176**
	lg_Crime density	0.190**	0.295**	0.156**

model that explained 73.0% of the variance in MHP (Supplementary Table 4). Although this model is not as good in terms of the predictive power as the model presented in Table 4b, it allows us to estimate the combined effects of greenspace proximity and SES/ethnicity-race factors on mental health conditions at the census-tract level.

3.4 Model performance in different sections of the city

As stated earlier, the three subsections of the City of San Diego (Figure 1) have very different socioeconomic and demographic characteristics [41]. Descriptive statistics of all the variables used in the analysis are presented in Supplementary Table 4. Using the same set of independent variables in Table 4b, plus the greenspace and park accessibility and crime density, we ran stepwise regression for these three subsections of the city (Figure 1). The models explained 94.3% (R_{adj}^2 = 0.936), 93.8% (R_{adj}^2 = 0.932), and 85.4% (R_{adj}^2 = 0.845) of the variance in MHP in the South, Middle, and North subsections, respectively, and all three models are statistically significant with F values ranging from 97.549 to 149.971 (p-values < 0.001). Educational attainment, percent below poverty, hours worked, percent of Black population, and regular checkups are the variables that persistently entered the models for all three subsections (Table 6), suggesting robust influences of these factors on MHP spatially. The other variables, however, had somewhat different impacts. For example, percent of Hispanic population is positively related to MHP in the southern and northern subsections of the city, but not for the middle subsection. The same is true for the housing affordability. Such changes in the variables included in the models of the different subsections of the city suggest that in different areas, the overall mental health conditions may be determined by different factors. Although a single general model structure can be determined, it may be difficult to produce a single predictive empirical model that can be applied to all regions.

4 Discussion

4.1 Impact of greenspace accessibility on MHP

Greenspace is a vital neighborhood resource and the relationship between greenspace and a range of beneficial macro social and health outcomes is well documented. While the exact causal paths between greenspace and social and health outcomes remain an open question, this study highlights important factors along this path that contributes to our understanding of the relationship between greenspace and mental health.

While it is important to distinguish between accessibility and utilization (proximity may not necessarily equate to utilization), findings from this study suggests that accessibility, measured as distance or proximity to greenspace and parks, is an important precursor to utilization. We find that factors known to be associated with health disparities, such as living in poverty or being identified as a member of an ethnic or racial minority (Hispanic

Table 6: Results of stepwise regression for the three subsections of the city. All variables are log-transformed. Census tracts with population less than 200 are removed from analysis. Column "B" is the regression coefficients and "Beta" is the standardized regression coefficients.

Section		B	Std. error	Beta	t	Sig.	Tolerance	VIF
South	(Constant)	3.409	0.662		5.148	0.000		
	lg_Poverty	0.091	0.022	0.268	4.073	0.000	0.231	4.324
	lg_Hispanic	0.113	0.022	0.294	5.183	0.000	0.310	3.221
	lg_Degree	-0.097	0.020	-0.280	-4.934	0.000	0.310	3.227
	lg_Checkup	-0.926	0.232	-0.205	-3.989	0.000	0.381	2.628
	lg_Black	0.027	0.007	0.160	3.988	0.000	0.619	1.614
	lg_Affordability	0.074	0.027	0.143	2.760	0.008	0.371	2.695
	lg_Hours worked	-0.540	0.218	-0.120	-2.472	0.016	0.425	2.354
Middle	(Constant)	4.481	0.463		9.670	0.000		
	lg_Poverty	0.128	0.023	0.370	5.500	0.000	0.230	4.348
	lg_Degree	-0.166	0.019	-0.422	-8.734	0.000	0.446	2.243
	lg_Hours worked	-0.508	0.138	-0.194	-3.686	0.000	0.376	2.661
	lg_Checkup	-1.323	0.254	-0.226	-5.205	0.000	0.554	1.806
	lg_Black	0.028	0.009	0.127	3.043	0.003	0.599	1.670
	lg_Crime density	-0.041	0.017	-0.119	-2.392	0.020	0.424	2.357
North	(Constant)	4.098	0.334		12.265	0.000		
	lg_Hispanic	0.056	0.018	0.179	3.055	0.003	0.364	2.748
	lg_Poverty	0.033	0.009	0.180	3.785	0.000	0.555	1.802
	lg_Checkup	-1.307	0.136	-0.461	-9.638	0.000	0.546	1.832
	lg_Degree	-0.200	0.035	-0.337	-5.759	0.000	0.364	2.745
	lg_Hours worked	-0.319	0.106	-0.118	-3.022	0.003	0.815	1.227
	lg_Affordability	0.064	0.021	0.125	3.120	0.002	0.777	1.287
	lg_Black	0.015	0.006	0.108	2.593	0.011	0.719	1.392

or Black), are also associated with both higher prevalence of poor mental health and greater distance to greenspace and parks. Second, sociodemographic characteristics such as race and poverty, continue to be important predictors of MHP beyond accessibility to greenspace. That is, our regression results suggest that sociodemographic factors such as living in poverty, Hispanic communities, and low educational attainment are more important than accessibility to greenspace in explaining MHP. In short, greenspace accessibility, while important when being considered alone, does not provide much additional explanation to the variation in MHP beyond what sociodemographic variables already have. This supports the continued need to consider mental health related policies through a range of interventions that address poverty and challenges that come with identifying as a member of an underprivileged racial and ethnic minority.

In summary, our initial correlation analysis shows that the longer the distances to greenspace, tree cover, and parks, the poorer the mental health conditions. This conforms to prior studies at the national level by comparing cities with different amount of urban greenspace [43,44]. Additionally, our results suggests that areas that have higher SES had greater accessibility to greenspace, tree cover, and parks, which supports previous studies as well [43,45]. Overall, census tracts with higher income, educational attainment, full-time employment, proportion of White residents, rates of regular checkups, and

greater accessibility to any measure of greenspace had better mental health conditions, which reflects well-known health disparities in the US [2,7]. This also suggests that the City of San Diego follows generally the same patterns as many other cities in the US related to health disparities associated with sociodemographic and other variables. For example, studies have found that lower rates of income and educational attainment are linked to higher rates of mental illness [36,46,47] similar to what our results show. Additionally, it has been found that Black and Hispanic populations had poorer health outcomes in general in the US [48].

When constructing our regression model using the SES and race-ethnicity variables against MHP, the primary relationship of interest between proximity to greenspace and mental health did not reach statistical significance, which suggests that the effects of greenspace accessibility on mental health conditions are most likely to be indirect. For both the overall model (Table 4b) and the models for different subsections of the city (Table 6), a core set of socioeconomic and race-ethnicity variables appear to be consistent in predicting MHP. Specifically, percent below poverty and Hispanic populations have a consistently positive relationship with MHP while education has a consistently negative relationship with MHP across all the models. In addition, household income and regular checkups — important independent variables in the MHP models (Table 4b) — were all correlated in some fashion to the accessibility to greenspace or parks (Table 3). The partial correlation analysis confirms our conclusion that the effects of greenspace and parks on MHP can be substituted by a selected group of SES and race-ethnicity variables (Table 5). Previous studies have presented similar results. For example, in Kabisch's [49] systematic review of 25 papers, although there was some evidence of a positive effect of greenspace on mental health, many of these studies found that socioeconomic and socio-demographic confounders, such as median household income, had the greatest impact on health outcomes. On the other hand, we were able to combine the greenspace proximity variables with a subset of the SES variables to build a model with relatively high predictive power (Supplementary Table 4), which will allow us to estimate the combined effects of greenspace proximity and SES/race-ethnicity factors on mental health.

While the unique distribution pattern of vegetation and parks relative to population distribution in the City of San Diego could be an important factor, the overall conclusion that the MHP is more influenced by SES and race-ethnicity factors than the physical settings of different communities is consistent with prior studies [9]. The models constructed in this study demonstrate how policymakers, community leaders and program providers can advocate for policies and implement interventions at a local level to make communities healthier. For example, policies should not only focus on increasing the total area of greenspace but must also be accompanied by resources for interventions to encourage preventative behaviors (i.e., checkups), or improve the income, education, and employment opportunities of underserved communities. Also critical are policies and interventions that target health disparities related to race and ethnicity, as studies have shown that there are disparities in mental healthcare for minorities. These policies may include increasing culturally competent healthcare measures to increase health equity in San Diego [50]; or diversifying the mental health workforce and improving provider and patient education [51].

4.2 Regression models

The models presented in Table 4b explained more than 93% of the variance in the dependent variable MHP, which is uncommon for socioeconomic/demographic data analyses. However, the model is statistically sound and we carefully addressed the issue of collinearity among the independent variables in the modeling process (Supplementary Part 2). One possible explanation is that the independent variables we used in the model likely replicated the variables that the CDC used to estimate the census-tract MHP values. Although we do not have the details of the CDC models, the description of their modeling process seemed to suggest that similar variables were considered [23]. However, our purpose is not to validate the CDC's modeling process, but to use these models to specify the factors and relationships that can be used to provide meaningful information in resource allocation and preventive intervention procedures. For example, one of our models (Supplementary Table 2, Part a) suggests that the most effective way to reduce MHP might be decreasing the percentage of uninsured population that itself is a function of SES and race-ethnicity factors (Supplementary Table 1), which calls for health insurance programs to cover the entire population. Additionally, promotion of regular checkups and raising the level of educational attainment might be effective interventions to reduce issues of poor mental health as these variables entered the regression models consistently. We also found that in different subsections of San Diego, different factors should be considered in resource allocation for managing MHP beside the common factors of the percent of unin-

sured population and regular checkups. For example, in both the southern and northern subsections, attention should be paid to census tracts with relatively high percentages of Hispanic and Black population to provide services to treat mental health related issues, but in the middle subsection, the focus should be shifted to census tracts with large Black population (Table 6). Similarly, housing affordability is an important factor of poor mental health, but not so in the middle section.

4.3 Limitations and possible future investigations

First and foremost, we took an empirical modeling approach in exploring the relationships between mental health conditions and potential factors including greenspace proximity and SES and race-ethnicity variables at the census tract-level. Even though we based our methodology on findings from previous studies and we are confident that our results are meaningful in discussions of possible policy and/or intervention measures to improve mental health conditions for communities in San Diego, there are inherent shortcomings of the empirical approach when compared with other modeling studies with strong theoretical basis, which may limit our capability to fully separate the effects of these interacting factors on mental health conditions in the analysis. For example, results from surveys and interviews can link individuals' mental health conditions directly to the stressors. As the individual-level data are aggregated into increasingly larger spatial units (e.g., census blocks, blockgroups, and census tracts) there is the risk that the original relationships may be obscured due to the "modifiable areal unit problem" (MAUP) commonly seen in census or any spatially aggregated data [52–54], which may lead to the "ecological fallacy" when the correlations calculated using the census-tract data may not represent the individual relationships within these census units [55,56]. Therefore, the empirical models based on the census-tract data may mask the individual level relationships. To guard against misinterpretation of our results, we specifically emphasized the logic of variable screening in regression modeling, took extra caution in the interpretation of the models and did not accept any results simply based on the statistical validity.

With this in mind, we would like to point out that the findings in this study are subject to several limitations. First, although the BRFSS are considered a gold-standard in telephone-based surveys [23,57], the mental health data is self-reported and may be misclassified because they are not objectively verifiable and are further subject to recall bias [22]. Although a 2013 meta-analysis that compared nationwide BRFSS data confirmed that the BRFSS mental health measures had moderate to excellent reliability [58], an external validation study comparing the 500 Cities estimates to Boston BRFSS direct survey data found that the 500 Cities mental health measures were not within the 95% confidence intervals of the direct survey estimates and most likely overestimated the prevalence of mental health [59]. Generally, it is important to understand that poor mental health is an estimate and could be subject to error and bias.

Second, although our data suggests that the occurrence of poor mental health conditions in San Diego is correlated to both socioeconomic and racial disparities, we could not specify the actual mechanisms that are creating these apparent linkages. It should be noted that the relationships between mental health conditions and SES, race-ethnicity, and accessibility to greenspace identified in this study are based on summaries or average values at the census tract scale, rather than for individuals, which may have obscured certain mechanisms for such linkages at the level of personal experiences. The SAE data "represent the expected prevalence estimate of a health indicator for census tracts and cities given the socio-demographic characteristics (age, gender, and race/ethnicity) of the individuals in the census tract or city, census tract-level poverty, and unmeasured contextual factors in the county and state" [23]. It is important to understand that the specific values do not represent the actual observed prevalence of health indicators in the census tracts or cities [23]. On the other hand, our results indeed suggest that by looking at the data at the census-tract level, there were statistically significant and logical links between the SES and race-ethnicity variables we considered and the mental health conditions. Therefore, our models should be useful in predicting spatial patterns of mental health conditions in San Diego and in providing the basis for allocation of resources and services accordingly. These results may also be useful in formulating community-level policies of intervention and mitigation for potential mental health-related issues. However, to further investigate the causes of poor mental health conditions, it is necessary to conduct interviews or surveys with individuals within these census tracts of interest to understand the relationships between mental health conditions and different types of stressors. This would allow us to look more directly at the mechanisms that are at play.

Finally, even though we considered both the quantity (total greenspace by vegetation types) and quality (tree canopy cover) of urban greenspace, proximity does

not guarantee actual usage of greenspace. Our proximity measures are less sensitive to whether a patch of greenspace is located within or outside a given census tract as a measure of accessibility. Nevertheless, we could not measure the actual greenspace utilization behaviors by residents within or outside the census tracts. Nor did we consider the possibility of an adjacent census tract with a higher percentage of greenspace than the others in the surrounding, allowing for the possibility of residents living in the greenspace barren tracts crossing over to the adjacent tract to use greenspace. All these are important questions regarding designing and implementing intervention policies and will need to be answered using individual utilization surveys.

5 Conclusion

In this study, we use the CDC 500 Cities Project public health dataset to evaluate the possible associations between MHP and accessibility to greenspace in the City of San Diego. Our study suggests that there are associations between the greenspace proximity and mental health conditions with statistically significant correlations (0.480 with mean distance to greenspace, 0.342 with distance to locations with 30–69% tree canopy cover, and 0.234 with distance to parks, all at p-value < 0.01). These relationships indicate that the longer the distances to greenspace and parks within a given census tract in San Diego, the poorer the overall mental health conditions. A regression model using only the greenspace proximity variables and crime density explained 37.8% of the variance in MHP. However, these associations are weak in comparison to the relationships between mental health conditions and SES and race-ethnicity factors (e.g., 0.916 with percent of uninsured population and -0.886 with median household income) in the City of San Diego. Specifically, higher SES was associated with better mental health conditions, and higher Hispanic and Black population percentages were associated with poorer mental health conditions as represented by statistically significant positive correlations (0.842 and 0.578 with percentages of Hispanic and Black population, respectively).

Using regression analysis, we built a model using the SES and race-ethnicity variables to predict the variation of mental health conditions across the City of San Diego. We also constructed models for three subsections of the city with very different socioeconomic and demographic characteristics. However, in both the all-city model and the models of subsections, the greenspace proximity variables did not enter the models with statistical significance. Partial correlation analysis confirmed that the effects of greenspace proximity, as indicated by statistically significant positive correlations with MHP, can be fully represented by a selected group of SES and race-ethnicity variables, such as educational attainment, percent of uninsured, and percent of Hispanic population. On the other hand, it is possible to build empirical models that combine both greenspace proximity and SES and race-ethnicity variables by carefully screening SES and race-ethnicity variables that do not overlap with the greenspace proximity variables' effects on MHP.

The results of this study can inform policymakers, public health officials, community leaders, and program providers about the relationships between mental health conditions, socioeconomic-demographic factors, and greenspace proximity. To the best of our knowledge, this study is probably one of only a few that focused on the City of San Diego, the second largest city in the State of California. With results that are consistent with studies conducted in other metropolitan areas, this study contributes to the body of literature on the socioeconomic and spatial factors related to mental health. Specially, this study highlights the idea that policies to increase the square area of greenspace in an urban area are necessary, but probably not sufficient to promote improved mental health conditions. Also critical are policies and interventions that promote preventative behaviors and improve the income, education, and employment opportunities of residents living in underserved communities.

Abbreviations

ACS	American Community Survey
ARJIS	Automated Regional Justice Information System
BRFSS	Behavioral Risk Factor Surveillance System
CDC	Centers for Disease Control and Prevention
GIS	Geographic Information System
MHP	Mental Health Prevalence
NDVI	Normalized Difference Vegetation Index
SAE	Small Area Estimates
SANDAG	The San Diego Association of Governments
SANGIS	The San Diego Geographic Information Source
SES	Socioeconomic Status
VIF	Variance Inflation Factor

Acknowledgments: The authors thank Dr. Suzanné Walther for assisting the lead author in the initial analysis of the CDC data, Dr. Adina Batnitzky for her suggestion to

incorporate crime data in the analysis, and Dr. Susannah Stern for her support to the research process in general, all from the University of San Diego. Thanks to two anonymous reviewers, whose suggestions and comments have significantly improved the manuscript from its earlier version. The authors are also grateful for the support and understanding from the journal editors during the review and revision processes.

Funding information: The authors state no funding involved.

Author contributions: AW and ZY designed the project and methodology; AW procured datasets and performed data processing; AW and ZY conducted data analysis and interpreted results; ML provided critical contextual information and assisted in interpretation of results; AW wrote the first draft, and ML and ZY assisted in revision and editing.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: All data used in this study are from publicly accessible sources and fully described in the data section. The 180-day crime data covering the time period prior to October 2019 are available as a shapefile (compressed) attached to the Supplementary Information or provided electronically by the corresponding author upon request.

Supplementary materials: The online version of this article contains supplementary materials available at https://doi.org/10.1515/openhe-2021-0004.

References

[1] US Centers for Disease Control and Prevention (CDC). Mental Health: US Centers for Disease Control and Prevention; 2021 [updated June 28 2021]. Available from: https://www.cdc.gov/mentalhealth/learn/index.htm.

[2] Office of the Surgeon General (US), Center for Mental Health Services (US), National Institute of Mental Health (US). Mental Health: Culture, Race, and Ethnicity: A Supplement to Mental Health: A Report of the Surgeon General. Rockville (MD): Substance Abuse and Mental Health Services Administration (US); 2001 Aug.

[3] Nutsford D, Pearson AL, Kingham S. An ecological study investigating the association between access to urban green space and mental health. Public Health. 2013;127(11):1005–11. doi: 10.1016/j.puhe.2013.08.016.

[4] Twohig-Bennett C, Jones A. The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes. Environ Res. 2018;166:628–37. doi: 10.1016/j.envres.2018.06.030.

[5] Kondo MC, Fluehr JM, McKeon T, Branas CC. Urban Green Space and Its Impact on Human Health. Int J Environ Res Public Health. 2018;15(3):445. doi: 10.3390/ijerph15030445.

[6] Collins RM, Spake R, Brown KA, Ogutu BO, Smith D, Eigenbrod F. A systematic map of research exploring the effect of greenspace on mental health. Landsc Urban Plan. 2020;201:103823. doi: 10.1016/j.landurbplan.2020.103823.

[7] Browning M, Rigolon A. Do Income, Race and Ethnicity, and Sprawl Influence the Greenspace-Human Health Link in City-Level Analyses? Findings from 496 Cities in the United States. Int J Environ Res Public Health. 2018;15(7):1541. doi: 10.3390/ijerph15071541.

[8] Dzhambov A, Hartig T, Markevych I, Tilov B, Dimitrova D. Urban residential greenspace and mental health in youth: Different approaches to testing multiple pathways yield different conclusions. Environ Res. 2018;160:47–59. doi: 10.1016/j.envres.2017.09.015.

[9] Dzhambov AM, Browning M, Markevych I, Hartig T, Lercher P. Analytical approaches to testing pathways linking greenspace to health: A scoping review of the empirical literature. Environ Res. 2020;186:109613. doi: 10.1016/j.envres.2020.109613.

[10] City of San Diego. Population 2021. Available from: https://www.sandiego.gov/economic-development/sandiego/population.

[11] Taylor L, Hochuli DF. Defining greenspace: Multiple uses across multiple disciplines. Landsc Urban Plan. 2017;158:25–38. doi: 10.1016/j.landurbplan.2016.09.024.

[12] Mental Health & Function Green Cities: Good Health: University of Washington, College of the Environment; 2018 [updated August 16, 2018]. Available from: https://depts.washington.edu/hhwb/Thm_Mental.html.

[13] City of San Diego. Facts and Figures 2021. Available from: https://www.sandiego.gov/economic-development/sandiego/facts.

[14] UC San Diego School of Medicine. Underserved Population: UC San Diego School of Medicine; 2021. Available from: https://medschool.ucsd.edu/som/medicine/education/residency/combined-medicine-pediatrics/program/Pages/Underserved-Population.aspx.

[15] Wong TK, Sanchez M. Immigrant Integration in the City of San Diego. U.S. Immigration Policy Center (USIPC) at UC San Diego; 2020 Jun 29.

[16] Beyer KM, Kaltenbach A, Szabo A, Bogar S, Nieto FJ, Malecki KM. Exposure to neighborhood green space and mental health: evidence from the survey of the health of Wisconsin. Int J Environ Res Public Health. 2014;11(3):3453–72. doi: 10.3390/ijerph110303453.

[17] Astell-Burt T, Feng X. Association of Urban Green Space With Mental Health and General Health Among Adults in Australia. JAMA Netw Open. 2019;2(7):e198209. doi: 10.1001/jamanetworkopen.2019.8209.

[18] US Centers for Disease Control and Prevention (CDC). Health-Related Quality of Life (HRQOL): US Centers for Disease Control and Prevention; 2018. Available from: https://www.cdc.gov/hrqol/wellbeing.htm.

[19] Lanza K, Stone Jr B, Haardorfer R. How race, ethnicity, and income moderate the relationship between urban vegetation and physical activity in the United States. Prev Med. 2019;121:55–61. doi: 10.1016/j.ypmed.2019.01.022.
[20] United States Census Bureau. Explore Census Data n.d. Available from: https://data.census.gov/cedsci/.
[21] Polanski TM, Collins K, Walter S. San Diego Native Vegetation Mapping Project. 1985.
[22] [US Centers for Disease Control and Prevention (CDC). 500 Cities Project: 2016 to 2019; 2020. Available from: https://www.cdc.gov/places/about/500-cities-2016-2019/index.html.
[23] Kong AY, Zhang X. The Use of Small Area Estimates in Place-Based Health Research. Am J Public Health. 2020;110(6):829–32. doi: 10.2105/AJPH.2020.305611.
[24] Humphrey JL. Relevant social and spatial contexts for elementary school children: An examination of multiple scales. Health Place. 2016;40:178–84. doi: 10.1016/j.healthplace.2016.06.003.
[25] Bower KM, Thorpe Jr RJ, Rohde C, Gaskin DJ. The intersection of neighborhood racial segregation, poverty, and urbanicity and its impact on food store availability in the United States. Prev Med. 2014;58:33–9. doi: 10.1016/j.ypmed.2013.10.010.
[26] Ross NA, Tremblay SS, Graham K. Neighbourhood influences on health in Montreal, Canada. Soc Sci Med. 2004;59(7):1485–94. doi: 10.1016/j.socscimed.2004.01.016.
[27] Akpinar A, Barbosa-Leiker C, Brooks KR. Does green space matter? Exploring relationships between green space type and health indicators. Urban Forestry & Urban Greening. 2016;20:407–18. doi: 10.1016/j.ufug.2016.10.013.
[28] SANDAG, SANGIS. San Diego Open GIS Data Portal n.d. Available from: https://sdgis-sandag.opendata.arcgis.com/.
[29] United States Department of Agriculture USFS. 2016 Tree Canopy Cover for the National Land Cover Database: Production Data, Methods, Uses, and a Tour Through American Landscapes. 2017. Available from: https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/#table1.
[30] Kondo MC, Mueller N, Locke DH, Roman LA, Rojas-Rueda D, Schinasi LH, et al. Health impact assessment of Philadelphia's 2025 tree canopy cover goals. Lancet Planet Health. 2020;4(4):e149–e57. doi: 10.1016/s2542-5196(20)30058-9.
[31] Arcaya MC, Figueroa JF. Emerging Trends Could Exacerbate Health Inequities In The United States. Health Aff (Millwood). 2017;36(6):992–8. doi: 10.1377/hlthaff.2017.0011.
[32] Hellman DA, Naroff JL. The impact of crime on urban residential property values. Urban Studies. 1979;16:105–12. doi: 10.1080/713702454.
[33] Buonanno P, Montolio D. Housing prices and crime perception. Empir Econ. 2012;45:305–21. doi: 10.1007/s00181-012-0624-y.
[34] Fleming CM, Manning M, Ambrey CL. Crime, greenspace and life satisfaction: An evaluation of the New Zealand experience. Landsc Urban Plan. 2016;149:1–10. doi: 10.1016/j.landurbplan.2015.12.014.
[35] SANDAG. ARJIS n.d. Available from: https://www.sandag.org/index.asp?subclassid=21&fuseaction=home.subclasshome.
[36] Araya R, Lewis G, Rojas G, Fritsch R. Education and income: which is more important for mental health? J Epidemiol Community Health. 2003;57:501–5. doi: 10.1136/jech.57.7.501.
[37] Abdel-Samad M, Adams B, Williams M, DeConinck K. Community Perceptions of Affordable Housing in San Diego. The Lucky Duck Foundation; 2020.
[38] Hauke J, Kossowski T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. Quageo. 2011;30(2):87–93. doi: 10.2478/v10117-011-0021-1.
[39] Griffith DA, Amrhein CG. Multivariate Statistical Analysis for Geographers: Prentice Hall, Inc.; 1997. p. 345.
[40] Florida R, Mellander C. Segregated City: The Geography of Economic Segregation in America's Metros. Martin Property Institute, Rotman School of Management, University of Toronto; 2015.
[41] McArdle N. Race, Place, and Opportunity: Racial Change and Segregation in the San Diego Metropolitan Area: 1990 - 2000. UCLA: The Civil Rights Project / Proyecto Derechos Civiles. 2002.
[42] Johnson RJ. Multivariate Statistical Analysis in Geography. Longman: New York; 1980. p. 280.
[43] Wen M, Zhang X, Harris CD, Holt JB, Croft JB. Spatial disparities in the distribution of parks and green spaces in the USA. Ann Behav Med. 2013;45(Suppl 1):S18–27. doi: 10.1007/s12160-012-9426-x.
[44] Williams TG, Logan TM, Zuo CT, Liberman KD, Guikema SD. Parks and safety: a comparative study of green space access and inequity in five US cities. Landsc Urban Plan. 2020;201:103841. doi: 10.1016/j.landurbplan.2020.103841.
[45] de Vries S, Buijs AE, Snep RPH. Environmental Justice in The Netherlands: Presence and Quality of Greenspace Differ by Socioeconomic Status of Neighbourhoods. Sustainability. 2020;12(15):5889. doi: 10.3390/su12155889.
[46] Gresenz CR, Sturm R, Tang L. Income and Mental Health: Unraveling Community and Individual Level Relationships. J Ment Health Policy Econ. 2001;4:197–203.
[47] Macintyre A, Ferris D, Gonçalves B, Quinn N. What has economics got to do with it? The impact of socioeconomic factors on mental health and the case for collective action. Palgrave Commun. 2018;4(1):10. doi: 10.1057/s41599-018-0063-2.
[48] Dubay LC, Lebrun LA. Health, behavior, and health care disparities: disentangling the effects of income and race in the United States. Int J Health Serv. 2012;42(4):607–25. doi: 10.2190/HS.42.4.c.
[49] Kabisch N. The Influence of Socio-economic and Socio-demographic Factors in the Association Between Urban Green Space and Health. In: Marselle MR, Stadler J, Korn H, Irvine KN, Bonn A, editors. Biodiversity and Health in the Face of Climate Change. Cham: Springer International Publishing; 2019. p. 91–119.
[50] Kim-Godwin YS, Clarke PN, Barton L. A model for the delivery of culturally competent community care. J Adv Nurs. 2001;35(6):918–25. doi: 10.1046/j.1365-2648.2001.01929.x.
[51] McGuire TG, Miranda J. New evidence regarding racial and ethnic disparities in mental health: policy implications. Health Aff (Millwood). 2008;27(2):393–403. doi: 10.1377/hlthaff.27.2.393.
[52] Fotheringham AS, Wong DWS. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. Environment and

Planning A: Economy and Space. 1991;23(7):1025–44. doi: 10.1068/a231025.
[53] Horner MW, Murray AT. Excess Commuting and the Modifiable Areal Unit Problem. Urban Studies. 2002;39(1):131–9. doi: 10.1080/00420980220099113.
[54] Nelson JK, Brewer CA. Evaluating data stability in aggregation structures across spatial scales: revisiting the modifiable areal unit problem. Cartography and Geographic Information Science. 2017;44(1):35–50. doi: 10.1080/15230406.2015.1093431.
[55] Openshaw S. Ecological Fallacies and the Analysis of Areal Census Data. Environment and Planning A: Economy and Space. 1984;16(1):17–31. doi: 10.1068/a160017.
[56] Piantadosi S, Byar D, Green S. The Ecological Fallacy. Am J Epidemiol. 1988;127(5):893–904. doi: 10.1093/oxford-journals.aje.a114892.
[57] Fitzpatrick KM, Willis D. Chronic Disease, the Built Environment, and Unequal Health Risks in the 500 Largest U.S. Cities. Int J Environ Res Public Health. 2020;17(8):2961. doi: 10.3390/ijerph17082961.
[58] Pierannuzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. BMC Med Res Methodol. 2013;13(49):1–14. doi: 10.1186/1471-2288-13-49.
[59] Wang Y, Holt JB, Zhang X, Lu H, Shah SN, Dooley DP, et al. Comparison of Methods for Estimating Prevalence of Chronic Diseases and Health Behaviors for Small Geographic Areas: Boston Validation Study, 2013. Prev Chronic Dis. 2017;14:170281. doi: 10.5888/pcd14.170281.
[60] Earickson R, Harlin J. Geographic Measurement and Quantitative Analysis. New York: Macmillan College Publishing Co.; 1994. p. 350.
[61] County of San Diego Department of Planning and Land Use. Chaparral Vegetation Community. n.d. Available from: https://www.sandiegocounty.gov/content/dam/sdc/pds/mscp/docs/Biodiversity/handoutvegcomm3.pdf.