



Using Google Street View to examine associations between built environment characteristics and U.S. health outcomes

Quynh C. Nguyen^{a,*}, Sahil Khanna^b, Pallavi Dwivedi^a, Dina Huang^a, Yuru Huang^a, Tolga Tasdizen^c, Kimberly D. Brunisholz^d, Feifei Li^e, Wyatt Gorman^f, Thu T. Nguyen^g, Chengsheng Jiang^h

^a Department of Epidemiology and Biostatistics, University of Maryland School of Public Health, College Park, MD, United States

^b Master's in Telecommunications Program, University of Maryland, College Park, MD, United States

^c Department of Electrical and Computer Engineering & Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT, United States

^d Healthcare Delivery Institute, Intermountain Healthcare, Salt Lake City, UT, United States

^e School of Computing, University of Utah, Salt Lake City, UT, United States

^f Google Cloud, NY, New York, United States

^g Department of Epidemiology and Biostatistics, University of California San Francisco School of Medicine, San Francisco, United States

^h Maryland Institute for Applied Environmental Health (MIAEH), University of Maryland, College Park, MD, United States

ARTICLE INFO

Keywords:

Neighborhood
Built environment
Google Street View
Computer vision systems
Geographic information system
Health
Rural

ABSTRACT

Neighborhood attributes have been shown to influence health, but advances in neighborhood research has been constrained by the lack of neighborhood data for many geographical areas and few neighborhood studies examine features of nonmetropolitan locations. We leveraged a massive source of Google Street View (GSV) images and computer vision to automatically characterize national neighborhood built environments. Using road network data and Google Street View API, from December 15, 2017–May 14, 2018 we retrieved over 16 million GSV images of street intersections across the United States. Computer vision was applied to label each image. We implemented regression models to estimate associations between built environments and county health outcomes, controlling for county-level demographics, economics, and population density. At the county level, greater presence of highways was related to lower chronic diseases and premature mortality. Areas characterized by street view images as ‘rural’ (having limited infrastructure) had higher obesity, diabetes, fair/poor self-rated health, premature mortality, physical distress, physical inactivity and teen birth rates but lower rates of excessive drinking. Analyses at the census tract level for 500 cities revealed similar adverse associations as was seen at the county level for neighborhood indicators of less urban development. Possible mechanisms include the greater abundance of services and facilities found in more developed areas with roads, enabling access to places and resources for promoting health. GSV images represents an underutilized resource for building national data on neighborhoods and examining the influence of built environments on community health outcomes across the United States.

1. Introduction

Neighborhood environments can influence the ability of individuals and families to access necessary resources for achieving and maintaining good health. Neighborhood attributes have been linked with a broad array of health outcomes including mortality, (Wing et al., 1992; Tyroler et al., 1993; Morris et al., 1996; Eames et al., 1993; Townsend et al., 1988) life expectancy, (Clarke et al., 2010) mental health, (Truong and Ma, 2006) self-rated health, obesity, (Mujahid et al., 2008; Black

et al., 2010; Heinrich et al., 2008) and diabetes. (Lysy et al., 2013; Grigsby-Toussaint et al., 2010) Neighborhood built environments with mixed land use (residential, commercial uses, institutional) (Frank et al., 2004) may promote health because they position amenities and community resources where people live. Infrastructure like roads is critical because these connect people to goods, services and social networks.

Research on built environment characteristics can be expensive and time consuming. (Rundle et al., 2011) Neighborhood audits of built environmental features have typically entailed onsite visits and due to

* Corresponding author at: Department of Epidemiology and Biostatistics, University of Maryland School of Public Health, 2234B SPH Building, 4200 Valley Dr, College Park, MD 20742, United States.

E-mail address: qtnguyen@umd.edu (Q.C. Nguyen).

<https://doi.org/10.1016/j.pmedr.2019.100859>

Received 2 March 2019; Accepted 28 March 2019

Available online 09 April 2019

2211-3355/ © 2019 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

the cost and logistical challenges of these methods which include travel time and staff training, data on built-environment features are often limited in scale to a few neighborhoods or regions.(Bader et al., 2017) However, understanding the potential impacts of neighborhood design on health outcomes necessitates the inclusion of diverse, heterogeneous neighborhoods. Using road network data and Google Street View (GSV) images, we were able to construct neighborhood characteristics for geographically diverse areas across the entire United States. Computer vision tools were used to automatically process street segments—which dramatically lowered costs while offering new data resources for neighborhood research.

In recent years, public health and social scientists have started to utilize GSV to conduct innovative research, and those nascent studies suggest that GSV is a reliable and cost-effective tool.(Rundle et al., 2011; Naik et al., 2014; Kelly et al., 2012) Rundle and colleagues compared field audits of neighborhood features to GSV data and they found high levels of concordance especially for measures of pedestrian safety, traffic, and infrastructure for active travel. Small items or features that had temporal variability displayed lower levels of concordance. Another team utilized GSV to audit built environments in Indianapolis and St. Louis and found high inter-rater reliability.(Kelly et al., 2012) Outside the U.S., Silva and team found GSV to be a reliable and valid tool compared to in-person audits for assessing obesogenic built environment features in a heterogeneous urban area in São Paulo.(Silva et al., 2015) A group of European scientists utilized GSV to measure physical environment characteristics in London, Paris, Budapest and other cities.(Feuillet et al., 2016) Using computer vision models on approximately 1 million GSV images, Naik and colleagues created high resolution maps of perceived safety for 21 cities across the United States.(Naik et al., 2014)

Nevertheless, rural areas in particular have been understudied—although they make up over 97% of the land area in the United States.(U.S. Census Bureau, 2015) Rural areas have comparatively less access to some amenities for maintaining and promoting health such as health care resources, supermarkets, transit systems and nearby schools, recreational facilities, and cultural attractions within walking distance that encourage physical activity (Khan et al., 2009; Meit et al., 2014). Additionally, good road structures are important for providing access to jobs, facilitating the movement of goods and people, accessing health care and education, and providing links to social services. Rural roads may lack capacity, fail to provide needed connectivity to communities, and inadequately support freight travel.(TRIP, 2017) These features may help explain stark health disparities seen between rural and urban areas, with rural areas having much higher mortality, morbidity and chronic diseases.(Wilcox et al., 2000; Befort Christie et al., 2012; Eberhardt and Pamuk, 2004; Hartley, 2004; Parks et al., 2003; Eberhardt et al., 2001) Obesity, ischemic heart disease, chronic obstructive pulmonary diseases, limitation of activity due to chronic health conditions, leisure time physical inactivity, and mental illness are higher in rural counties than in urban or suburban counties.(Meit et al., 2014)

Since their launch in 2007, GSV has captured 20 petabytes of data, equivalent to 5 million miles of road around the world.(Farber, 2012) GSV images provide a unique lens into the local built environment with ground-level views not possible with other data sources such as satellite data. Street View image data also provide flexibility in allowing investigators to extract a variety of built environment features from one data source. Additionally, the geocoordinates associated with each image allows the use of flexible neighborhood boundaries to summarize built environment characteristics at different levels of aggregation (zip code, census tract, county, state). Other neighborhood data sources such as the U.S. Census provide complementary information on demographics and economic characteristics of residents.

Study Aims.

In this study, we leverage millions of GSV images and computer vision to create indicators of urban development based upon the physical features of the environment. We focus on the absence of

infrastructure and facilities because we believe that one main mechanism driving urban-rural disparities is fewer community resources and services found in rural communities. We examine whether our constructed indicators of urban development predict county level chronic disease, premature mortality, self-rated health, and health behaviors—controlling for population density as well as county demographic and economic characteristics. These diverse outcomes were selected in order to investigate the degree to which different dimensions of health were associated with our GSV measures. Previous research have linked neighborhood conditions to health behaviors (Saelens and Handy, 2008; Sallis et al., 2018), chronic conditions (Alvarado, 2016; Barrientos-Gutierrez et al., 2017), mental health (Galea et al., 2005; Evans, 2003) as well as mortality.(Hembree et al., 2005; Hankey et al., 2011)

2. Methods

2.1. Street view image collection

Using national road network data, we built a database of latitude and longitude coordinates representing all the street intersections in the United States. We focused on sampling images from street intersections in order to create a dataset that characterizes environments where people inhabit. In the United States, there are vast, sparsely populated roadless areas, especially mountain ranges and deserts. The roadway network files were accessed from the 2017 Census Topologically Integrated Geographic Encoding and Referencing data set. We downloaded all road types. We identified street intersections using PostgreSQL (an open-source object-relational database system) with the PostGIS plugin. The plugin is spatial database extender and enables location queries to be run in SQL. More information about the plugin can be found at <https://postgis.net/>.

Using these latitude and longitude coordinates, we retrieved GSV images for these locations. Between December 15, 2017-May 14, 2018, we used Google's Street View Image Application Programming Interface (API) to obtain images. Parameters for the API include the following: image size (640 × 640 pixels is the maximum image resolution for non-premium plan users), geographic location (geographic coordinates or addresses), field of view (zoom level), up or down angle of the camera relative to the Street View vehicle (default is 0), and heading (direction the camera is facing with 0 = north, 90 = east, 180 = south and 270 = west). Previously the API allowed users to download GSV images free of charge up to 25,000 map loads per 24-h period. However, on July 16, 2018, a new pay-as-you-go pricing plan went into effect for Maps, Routes, and Places. More information can be found at <https://developers.google.com/maps/documentation/streetview/usage-and-billing>.

We obtained four Street View images (directions: west, east, north and south) for each pair of coordinates to comprehensively capture 360 degree views of the environment. Image resolution was 640 × 640 pixels. We first sampled two-thirds of counties and then obtained all the intersections within the sampled counties. In total, we collected 16,171,605 images from a subset of 2143 counties in the United States. eFigure 1 displays the national coverage of our image data collection with sampling points dispersed across the United States. eTable 1 displays the number of images collected per state as well as the number of counties, by state, with image data.

2.1.1. Image data processing

We used Google's Vision API, Out-Of-Box, to label each Street View image. The API is able to detect thousands of different pre-defined items in images, ranging from wildlife to food to clothing. Users can output labels as .txt or .csv for further analysis. The API provides the advantage of having access to image classification algorithms that have been built using very large training sets. Computer vision is an established field and concepts employed by Google's API would also apply to other

software or specifically trained algorithms. Pricing information for the API can be found at the following website <https://cloud.google.com/vision/pricing>.

The API took less than one second to process each image and we utilized the API to analyze 16 million images from April 25–May 10, 2018. The API provides ten labels for each image. For this study, we focused on labels that characterized the built environment including 1) presence of highways (main road, especially connecting towns or cities), 2) rural area (sparsely spaced houses or buildings; limited surrounding infrastructure; unpaved roads), and 3) grassland (a large open area covered by grass, especially farmland used for grazing or pasture). More highways may represent more robust transit systems that enable the travel of goods and people. Conversely, more images labeled as rural area and grassland signal less urban development. Each image had a unique image identification number that was comprised of its latitude, longitude and camera direction. Image labels were merged with the images using this unique ID.

2.1.2. Quality control activities

In order to evaluate the accuracy of the computer vision API and considering time and participant fatigue, two coauthors manually labeled 300 images. Specifically, 50 random images each were selected from the following categories as determined by Google's computer vision API: highway = 1; highway = 0; grassland = 1; grassland = 0; rural area = 1; rural area = 0. Inter-rater reliability varied from 87% (rural area) to 94% (grassland) (eTable 2). Across the indicators, agreement between the manual labels and computer vision labels ranged from 82 to 95%. The number of manual labels is comparable to other GSV studies that have utilized 200–300 manually verified images and report similar agreement rates between human- and computer produced labels. (Hara et al., 2013; Movshovitz-Attias et al., 2015; Hyam, 2017)

2.2. County-level health outcomes

County health data were obtained from external sources that age-adjusted measures to the 2000 U.S. standard population. The most recent available data were obtained from the 2018 release of the County Health Rankings. Below we describe in more detail each of the health outcomes and their data sources. Data for *Years of Potential Life Lost* (YPLL) came from the National Vital Statistics System (2014–2016). YPLL is the years of potential life lost before age 75 presented per 100,000 population. Data on chronic conditions, self-rated health, and health behaviors were obtained from the 2014 Behavioral Risk Factor Surveillance System (BRFSS). *Adult obesity* was assessed by the percentage of the adult population (age 20 and older) that reported a body mass index (BMI) ≥ 30 kg/m². *Diabetes* was assessed via the question, “Has a doctor ever told you that you have diabetes? (for women, outside of pregnancy).” General *self-rated health* was categorized as fair or poor vs. excellent, very good, and good. *Frequent Mental Distress* was the percentage of adults who reported ≥ 14 days in response to the 2016 BRFSS question, “Now, thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?” *Frequent Physical Distress* was the percentage of adults who reported ≥ 14 days in response to the question, “Thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?”

Physical inactivity was assessed by the percentage of adults aged 20 and over reporting no leisure-time physical activity in the past month. *Excessive drinking* was defined as the percentage of adults reporting heavy drinking (drinking > 1 (women) or 2 (men) drinks per day on average) or binge drinking (consuming > 4 (women) or 5 (men) alcoholic beverages on a single occasion in the past 30 days). *Teen Births* was defined as the number of births per 1000 female population, ages 15–19 and data were drawn from the National Vital Statistics System (NVSS), 2010–2016.

2.3. Census tract-level health outcomes

Supplemental analyses were done at the census tract level to examine whether associations between built environment characteristics and health outcomes at the county level are similarly observed at finer levels of geographies such as census tracts. However, health data were only available for a proportion of census tracts compared to county health data which are available for all counties. Census tract health outcomes came from the Disease Control and Prevention's (CDC) 500 Cities Project. 2015 adult outcomes included obesity, diabetes, frequent physical distress, frequent mental distress, physical inactivity and binge drinking. We also examined limited access to healthy foods (% of the population living more than ½ mile from the nearest supermarket, supercenter, or large grocery store) and dental care (% aged ≥ 18 years who report having been to the dentist or dental clinic in the previous year).

2.4. Analytic approach

ArcGIS Desktop software (ESRI, Inc.) was used to create choropleth maps and the geographical data were obtained from the 2016 U.S. Census TIGER/Line Shapefiles. County and census travel level built environment characteristics were categorized into tertiles—high, moderate, and low using cut points that grouped one third of areas in the highest tertile, another third in the moderate tertile, and another third in the lowest tertile for each of the variables. Tertiles were chosen to ease interpretation of results and allow for non-linearities in the association between area characteristics and health outcomes. Area level health outcomes as defined above were modeled as continuous variables (e.g., county obesity rates). Models controlled for population density (population per square mile) and county sociodemographic characteristics. County-level demographic and economic characteristics were obtained from the 2010–2014 American Community Survey 5-year estimates and included the following: percent < 18 years old, percent 65 years and older, percent Hispanic, percent non-Hispanic black, percent non-Hispanic Asian, percent American Indian/Alaska Native, percent not proficient in English, economic disadvantage (standardized factor score summarizing the following four variables: percent unemployed; percent with some college, percent with high school diploma, percent children in poverty, and percent single parent households).

We implemented adjusted linear regression models to estimate differences in the prevalence of county health conditions (95% CI) by tertile of built environment characteristic, controlling for area compositional characteristics. The lowest tertile served as the referent group. Reported prevalence differences represent comparisons between the prevalence of health outcomes for those living in the 3rd tertile (vs. 1st tertile) and 2nd tertile (vs. 1st tertile) for area characteristics. Because our analyses are cross-sectional, the results of the linear regression models are prevalence differences (rather than risk differences as would be the case of longitudinal data). *Positive* prevalence differences indicate that individuals living in areas in the 3rd and 2nd tertile have *higher* prevalence of adverse health outcomes than those in the 1st tertile. *Negative* prevalence differences indicate that individuals living in areas in the 3rd and 2nd tertile have *lower* prevalence of adverse health outcomes than those in the 1st tertile. Models were run separately for each health outcome. Across models, sample size varied (2–5%) due to missing outcome or predictor variables. We evaluated statistical significance at $p < 0.05$ and reported robust standard errors. Data processing and statistical analysis tasks were performed with Stata MP15 (StataCorp LP, College Station, TX). The study was approved by the University of Maryland Institutional Review Board.

3. Results

Geographical distribution of the indicators, by county, and sample street view images are displayed in Figs. 1–3 (Geographical distribution

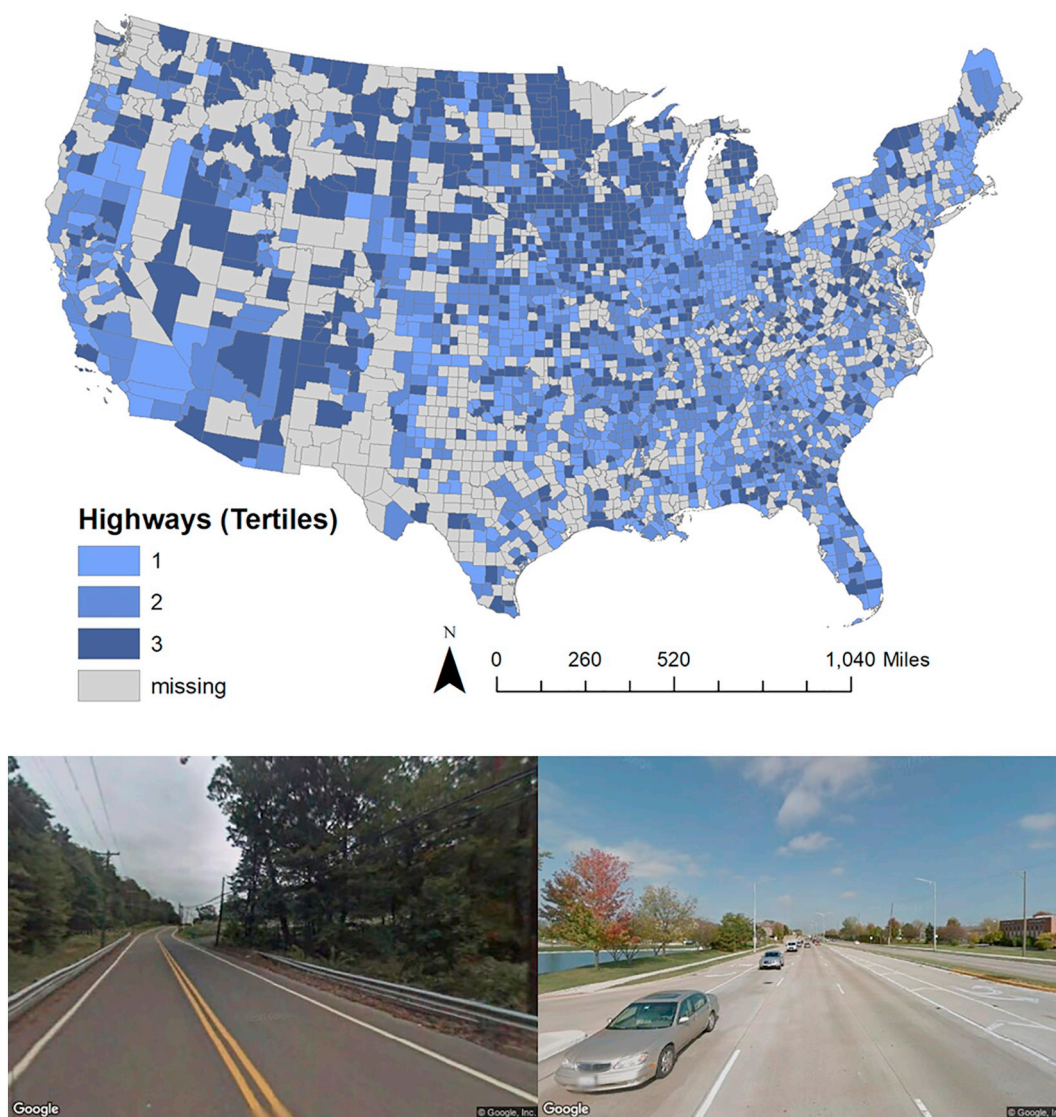


Fig. 1. Percent of street intersection images with highways, by county. Data source: Google Street View images.

of the three indicators, by state are displayed in eFigures 2–4). States with the most highways in street intersection images included Minnesota (28%), Nevada (21%) and Montana (21%). Places with the most street intersection images labeled as rural areas included Oklahoma (33%), Mississippi (29%), and Louisiana (25%). Grasslands were most prevalent in street intersection images captured from North Dakota (33%), South Dakota (25%), and Wyoming (23%).

Table 1 displays descriptive statistics for the 16.1 million Street View images covering 2143 counties across the United States, with representation from each of the fifty states including territories like Puerto Rico and Guam. Highways were detected in about 11% of images. Additionally, about 14% of images were labeled as rural areas (having limited infrastructure and buildings) and 5% of images were labeled as grasslands (a large open area covered by grass, especially farmland used for grazing or pasture). County-level summaries were created by averaging all the images pertaining to a given county. Correlations between the built environment indicators were generally low (range in |r values|: 0.09–0.25).

eTable 3 presents the results of adjusted linear regression analyses examining associations between population characteristics and GSV-derived built environment characteristics. Percent < 18 years old was related to more rural areas and more grasslands. Economic

disadvantage was related to fewer highways, more rural areas, and fewer grasslands. Greater population density was related to modestly fewer highways.

Table 2 and Table 3 display results of analyses relating county-level built environment predictors and county-level health outcomes—controlling for county level demographic and economic characteristics. Presence of highways was beneficial for outcomes (fair/poor self-rated health, diabetes, premature mortality, physical distress, mental distress, physical inactivity, and teen births) but was non-significant for obesity. For instance, counties with the most highways in images had 452 fewer years of potential life lost per 100,000 population compared to counties with the least highways. Additionally, counties with the most highways saw a 0.81% increase in excessive drinking rates compared with counties with the fewest highways. In additional analyses, we examined the relationship between highways and motor vehicle mortality. More highways were associated with increases in motor vehicle mortality (eTable 4).

While presence of highways was associated with better health outcomes, indicators of less development were associated with worse health with some exceptions. Counties with higher percentages of street view images denoting rural areas (having limited infrastructure and buildings) had worse health in terms of higher obesity, diabetes, fair/

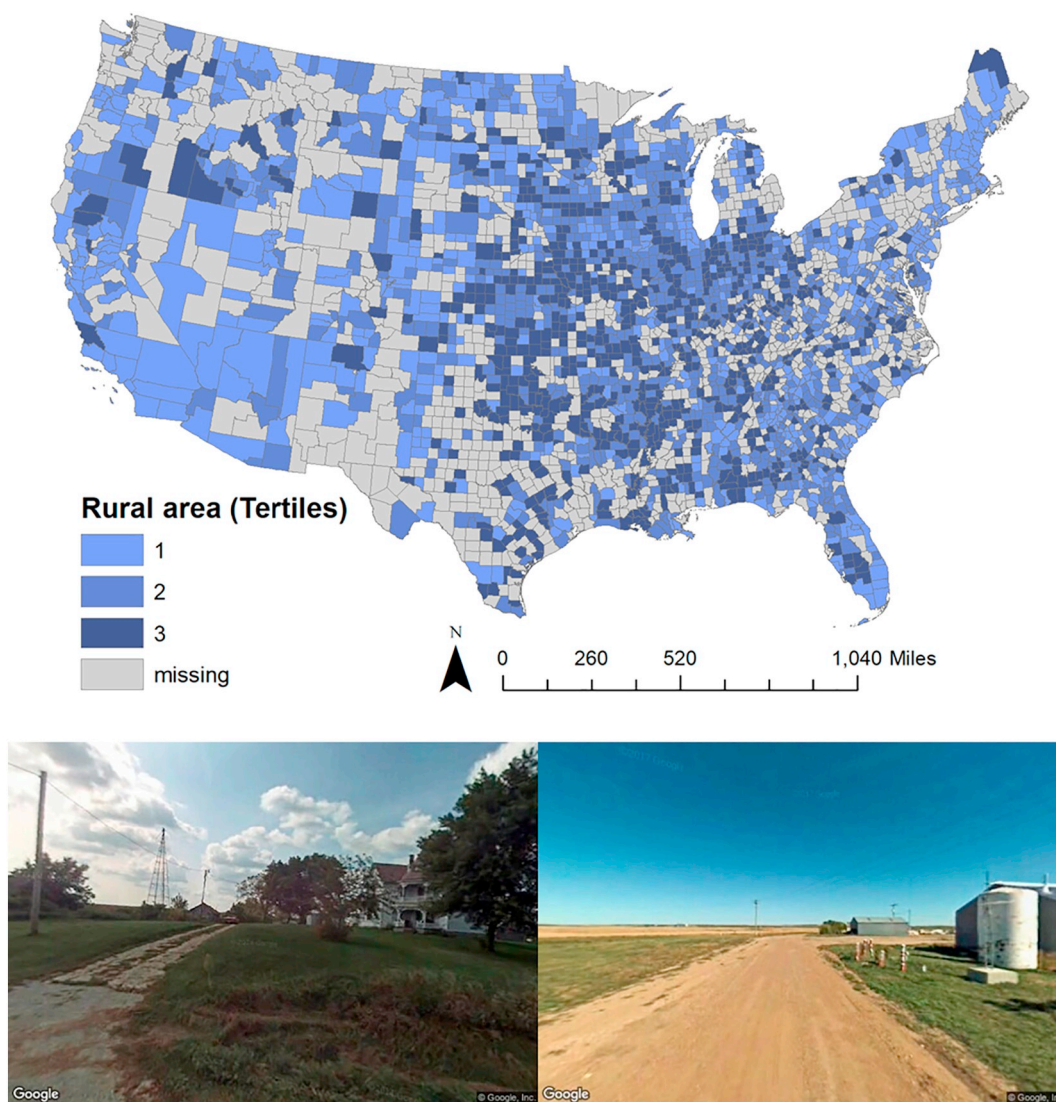


Fig. 2. Percent of street intersection images with rural area, by county. Data source: Google Street View images.

poor self-rated health, premature mortality, physical distress, physical inactivity and teen birth rates but had lower rates of excessive drinking. Counties with more grassland had higher obesity, physical inactivity and teen births but lower mental distress and physical distress.

Given that our GSV-derived indicator of rural area was consistently connected with worse outcomes, we implemented additional analyses to investigate possible mechanisms. Rural areas were correlated with fewer primary care physicians (per 100,000 population) and less access to recreational facilities at the county level (Table 4).

In further analyses, we examined health outcomes at the census tract level, controlling for compositional characteristics. Similar to associations seen at the county level—the GSV-derived measure of rural area was related to higher diabetes, physical distress, mental distress and physical inactivity but lower binge drinking. GSV-derived rural area was also associated with less access to healthy foods and dental care (Table 5).

Sensitivity analyses were also run comparing the GSV-derived variable for rural area and other indicators of rurality. A census tract was rural if the geographic centroid of the tract contained fewer than 2500 people. A GSV image was defined as rural if it displayed sparsely spaced houses or buildings; limited surrounding infrastructure; or unpaved roads. The GSV rural area variable was correlated with the census tract rural designation ($r = 0.58$) and with population density

($r = -0.70$). Table 5 displays these predictors in separate models and their relationships with census tract health outcomes. Tertiles of population density were similarly related to health outcomes but associations were smaller in magnitude than those seen for the GSV rural area indicator. Similar findings were also observed with the census-derived designation of rurality and the USDA rural-urban continuum codes (Table 5).

4. Discussion

The built environment has implications for health outcomes by structuring amenities, risks and resources. In this study, comparing across geographic areas in the United States, we found that indicators of greater area utilization and urban development were related to lower chronic disease prevalence, premature mortality, physical inactivity and teenage births. Possible mechanisms include the greater abundance of services and facilities found in more developed areas and the presence of major roads which is important for connecting people to places and each other, thereby enabling people to utilize resources for promoting health. Adverse associations were detected both at the county level and census tract level. Our models controlled for the socio-demographic composition of residents in an area. These results characterizing built environments by their level of infrastructure, may

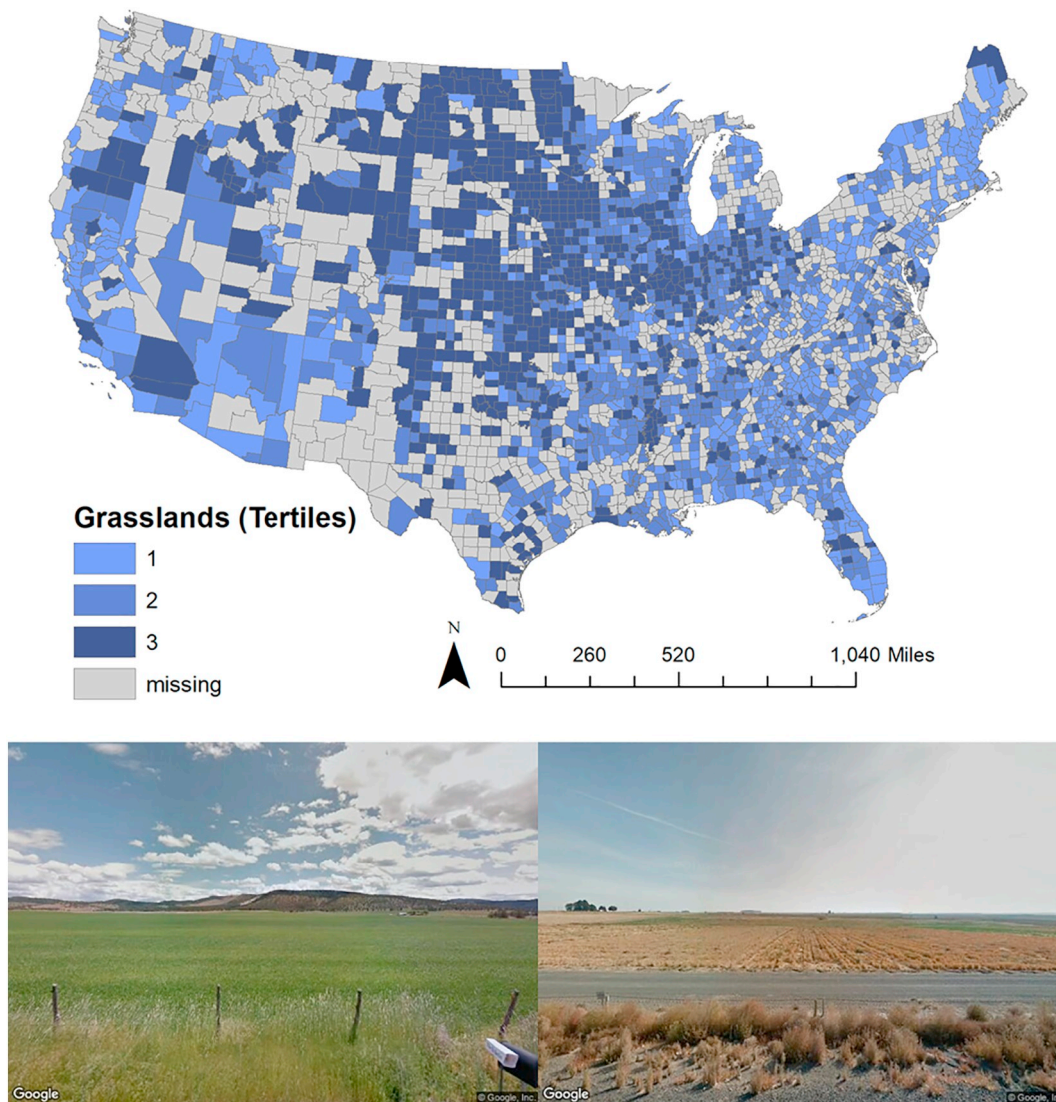


Fig. 3. Percent of street intersection images with grassland, by county. Data source: Google Street View images.

Table 1
Descriptive characteristics of Google Street View-derived built environment characteristics.

	Google Street View images		County summaries	
	N	Percent (standard deviation)	N	Percent (standard deviation)
Highway	16,172,373	11.36 (31.73)	2144	18.41 (14.31)
Rural area	16,172,373	14.23 (34.93)	2144	22.99 (16.95)
Grassland	16,172,373	5.49 (22.78)	2144	14.47 (18.23)

Neighborhood characteristics derived from street images collected between December 2017–April 2018 from Google’s Street View Image API.

indicate differential access to resources and services where people live and explain differential health outcomes.

Our study is unique in that it includes data from across the United States rather than a few select locations, which is more common for studies investigating environmental characteristics. We assessed national patterns and identified a robust pattern of health disparities in areas with less infrastructure. Study implications may include advocating for more health resources and structural investments in rural areas in order to mitigate against observed health disparities.

4.1. Study findings in context

Our study contributes to the nascent body of literature utilizing GSV to implement virtual neighborhood audits for neighborhood features such as walkability, (Brookfield and Tilley, 2016) physical disorder, (Mooney et al., 2014) retail alcohol stores, (Less et al., 2015) and urban greenery, (Li et al., 2015) Researchers have also implemented computerized approaches to label images for pedestrian count (Yin et al., 2015) and visual enclosure (i.e., proportion sky visible from a point on the street) (Yin and Wang, 2016)—measures that are connected with walkability. A previous study found Google’s Computer Vision API to be effective at characterizing the naturalness of urban areas with GSV images from the city of Edinburgh. (Hyam, 2017) In this paper, we extend the literature by scaling up to analyze millions of GSV images across the United States in order to examine the relationship between built characteristics and health outcomes.

Our GSV rural area indicator was associated with an array of adverse health outcomes and this is in alignment with research finding stark health disparities between rural and urban areas; for instance, lower physical activity (Parks et al., 2003) and higher obesity in rural areas. (Befort Christie et al., 2012) Research has found higher mixed-land use, street connectivity, and public transit to be positively

Table 2
Google Street View-derived predictors of county health outcomes^a.

	Percent with fair/poor health	Percent with diabetes	Percent with obesity	Years of potential life lost (per 100,000 people)	Percent with physical distress	Percent with mental distress
	Prevalence difference (95% CI) ^b	Prevalence difference (95% CI) ^b	Prevalence difference (95% CI) ^b	Prevalence difference (95% CI) ^b	Prevalence difference (95% CI) ^b	Prevalence difference (95% CI) ^b
Indicator of greater development						
Highway						
3rd tertile (highest)	-0.51 (-0.74, -0.29) [*]	-0.64 (-0.82, -0.46) [*]	-0.10 (-0.48, 0.29)	-452.07 (-626.65, -277.50) [*]	-0.27 (-0.40, -0.15) [*]	-0.36 (-0.47, -0.26) [*]
2nd tertile	-0.14 (-0.36, 0.07)	-0.23 (-0.40, -0.06) [*]	-0.07 (-0.42, 0.28)	-190.21 (-340.13, -40.30) [*]	-0.07 (-0.19, 0.05)	-0.13 (-0.23, -0.02) [*]
Indicators of less development						
Rural area						
3rd tertile (highest)	0.79 (0.55, 1.03) [*]	0.68 (0.50, 0.87) [*]	1.85 (1.44, 2.25) [*]	270.18 (77.13, 463.23) [*]	0.26 (0.13, 0.39)	0.10 (-0.02, 0.22)
2nd tertile	0.44 (0.20, 0.67) [*]	0.37 (0.18, 0.55) [*]	1.35 (0.95, 1.74) [*]	73.11 (-107.02, 253.25) [*]	0.13 (0.01, 0.26) [*]	-0.01 (-0.12, 0.10)
Grassland						
3rd tertile (highest)	0.10 (-0.14, 0.34)	-0.12 (-0.32, 0.08)	1.41 (0.99, 1.82) [*]	-24.69 (-202.35, 152.98)	-0.24 (-0.38, -0.11) [*]	-0.48 (-0.60, -0.36) [*]
2nd tertile	0.18 (-0.06, 0.42)	0.14 (-0.04, 0.33)	1.07 (0.69, 1.45) [*]	-38.24 (-201.87, 125.38)	-0.01 (-0.14, 0.12)	-0.12 (-0.23, -0.01)
N	2108	2108	2108	2074	2044	2044

^a County built environment characteristics categorized into tertiles with the lowest tertile serving as the referent group. Adjusted linear regression models were run for each predictor and outcome separately. Models controlled for county-level demographics: county-level demographics: county-level demographics: percent < 18 years old, percent 65 years and older, percent Hispanic, percent non-Hispanic black, percent non-Hispanic Asian, percent American Indian/Alaska Native, economic disadvantage, percent not proficient in English, and population density. Robust standard errors reported.

^{*} *P* < 0.05.

associated with meeting recommended physical activity guidelines and reductions in overweight/obesity.(Li et al., 2008) Additionally, people in rural areas may also have lower health care access due to increased travel distance and fewer health care providers.(Chan et al., 2007)

In our study, highways were associated with a reduction in most outcomes examined except for obesity. Previous literature has found the presence of highways correlated with restaurant density and in particular, fast food restaurants(Block et al., 2004; Chen et al., 2013)—which may negate some of the potential positive effects of greater infrastructure in a community. Research on proximity to roads has also suggested that they can expose individuals and families to harmful air pollutants, elevate the risk of respiratory and cardiovascular conditions, and increase noise disturbance, motor vehicle injuries and mortality.(Kim et al., 2004; Egan et al., 2003; Boothe and Shendell, 2008) Furthermore, the presence of highways, which may bring traffic and noise, may deter walking and other forms of physical activity. For people living in locations with adequate access to health resources, living away from major roads may be health beneficial, especially for those who have underlying conditions that can be aggravated by proximity to traffic. However, our analysis of street view data comparing (urban and rural) geographies across the United States suggests that individuals living in areas connected by highways experience a wide range of potential beneficial effects compared to those living in areas with fewer highways—these beneficial effects may be mediated by ease of travel to resources that highways provide.

4.2. Study strengths and limitations

The neighborhood built environment can promote health by locating neighborhood amenities or resources that are conducive for health or health behaviors. Thus far, investigations on built environment features in the U.S. have generally been limited to local studies because traditional neighborhood studies rely on people to perform neighborhood audits. Moreover, neighborhood studies in rural areas have been greatly under-investigated. In this study, we utilized road network data to build a national database of image search points for street intersections. The novel data collection strategy allowed us to capture street images from rural, suburban and urban areas, providing an extensive data source for future neighborhood research. We leveraged recently developed computer vision tools to produce area summaries of built environment characteristics. We then investigated the potential impact of neighborhood environments on chronic diseases and health behaviors.

Using GSV images offers three advantages that contribute to existing research on urban-rural health disparities. First, GSV images allow for assessments of built environment features, which complement other neighborhood data on population density and sociodemographic characteristics. Second, GSV indicators may offer more recent neighborhood data. Highly valuable and finely detailed neighborhood surveys like the Boston Neighborhood Survey (latest wave of data in 2010) are expensive and time-consuming, and thus difficult to update and conduct beyond a local geographical scale. Other indicators of rurality may also have a time lag of several years; the most recent USDA rural-urban continuum codes are available for 2010 at the census tract level and 2013 at the county level. A third advantage of GSV is that users have flexibility in creating neighborhood level summaries from GSV images that may be aggregated to any user-specified region (census tract, zip code, county, or other neighborhood-specific boundary).

Nonetheless, our study has limitations. We utilized proprietary software to conduct computer vision and generate pre-defined labels. As a result, we could not specify particular neighborhood indicators. To evaluate its performance, we compared with manual labels and found excellent levels of agreement. Of note, the built environment indicators selected for this study were large in size. The Google API may have lower accuracy for smaller objects(Hyam, 2017), as is the case for other computer vision tools.

Table 3
Google Street View-derived predictors of county behavioral health outcomes^a.

	Physical inactivity	Teen births	Excessive drinking
	Prevalence difference (95% CI) ^a	Prevalence difference (95% CI) ^a	Prevalence difference (95% CI) ^a
Indicator of greater development			
Highway			
3rd tertile (highest)	-0.99 (-1.41, -0.56)*	-2.20 (-3.19, -1.21)*	0.81 (0.54, 1.08)*
2nd tertile	-0.26 (-0.68, 0.15)	-0.54 (-1.52, 0.44)	0.14 (-0.10, 0.39)
Indicators of less development			
Rural area			
3rd tertile (highest)	2.57 (2.09, 3.05)*	2.88 (1.77, 4.00)*	-0.36 (-0.65, -0.06)*
2nd tertile	1.40 (0.95, 1.85)*	2.00 (0.92, 3.08)*	0.05 (-0.23, 0.33)
Grassland			
3rd tertile (highest)	1.47 (0.98, 1.95)*	1.19 (0.10, 2.28)*	0.28 (-0.01, 0.56)
2nd tertile	1.23 (0.78, 1.68)*	0.86 (-0.14, 1.86)	0.09 (-0.17, 0.36)
N	2108	2044	2108

^a County built environment characteristics categorized into tertiles with the lowest tertile serving as the referent group. Adjusted linear regression models were run for each predictor and outcome separately. Models controlled for county-level demographics: county-level demographics: percent < 18 years old, percent 65 years and older, percent Hispanic, percent non-Hispanic black, percent non-Hispanic Asian, percent American Indian/Alaska Native, economic disadvantage, percent not proficient in English, and population density. Robust standard errors reported.

* $p < 0.05$.

Table 4
Google Street View derived rural area (limited infrastructure) as a predictor of county health care access and exercise opportunities.

Rural area ^c	Primary care physician rate ^a	Exercise opportunities ^b
	Prevalence difference (95% CI) ^c	Prevalence difference (95% CI) ^c
3rd tertile (highest)	-13.96 (-17.89, -10.03)*	-9.39 (-11.73, -7.06)*
2nd tertile	-8.69 (-12.35, -5.03)*	-4.86 (-7.04, -2.69)*
N	2022	2108

^a Primary care physician = primary care physicians per 100,000 population, 2015.

^b Exercise opportunities = percent of the population with access to places for physical activity. Access was defined for urban census blocks as living within half a mile from a park or a mile from a recreational facility and defined for rural census blocks as living within 3 miles from a recreational facility, 2016.

^c County rural area indicator categorized into tertiles, with the lowest tertile serving as the referent group. Adjusted linear regression models were run for each predictor and outcome separately. Models controlled for county-level demographics: county-level demographics: percent < 18 years old, percent 65 years and older, percent Hispanic, percent non-Hispanic black, percent non-Hispanic Asian, percent American Indian/Alaska Native, economic disadvantage, percent not proficient in English, and population density. Robust standard errors reported.

* $p < 0.05$.

Moreover, Street View image data can only capture some of the neighborhood features within the community. For instance, image data does not allow for the creation of indicators for noise and perceived safety. Also, collected image data were from street intersections which offer unique viewpoints on local activity given that they are gathering points for traffic and people, but nevertheless, do not capture all important environmental features. As such, our measures are interpreted as the percentage of built environment features seen at these intersections.

Onsite field visits have enabled researchers to identify hundreds of neighborhood characteristics that impact health. Well-known neighborhood inventories include Irvine-Minnesota Inventory, (Day et al., 2006) the Pedestrian Environment Data Scan, (Clifton et al., 2007) and the Maryland Inventory of Urban Design Qualities. (Ewing et al., 2006) Utilizing computer vision may impact the type and depth of neighborhood features examined. In particular, computer vision models have difficulty with features that have variability across time or are small in

size. Understanding the context in which certain neighborhood features appear may also be difficult in virtual audits compared to onsite field visits. However, as this study demonstrates, computer vision models make possible national neighborhood studies incorporating millions of images.

An additional study limitation is that data on census tract level outcomes were only available for select cities and hence leaves out more rural areas of the country. Even among these select cities, we find that cities characterized by less infrastructure have worse health outcomes. While GSV and other technologies begin to enable larger-scale characterization of U.S. neighborhoods, data availability for geotagged health outcomes across wide areas of the United States continues to be an issue for neighborhood research.

Another study limitation is a possible temporal mismatch between health outcomes and GSV data. Google Street View API provides the most recent image available for a location. However, areas differ with regard to the rate at which their GSV image are updated. In our dataset, image dates ranged from 2007 to 2017; the median year was 2013. The main health outcomes were assessed between 2014 and 2016.

Moreover, the observational nature of the study inhibits causal inference; relationships reported here are observed associations rather than causal effects. Causal effects are difficult to estimate for neighborhood characteristics in particular because people are often not randomly assigned their residential environments (indeed previous and existing policies have led to high levels of residential segregation in certain communities). Further research with additional study designs, for example, involving changes in neighborhood conditions and changes in health outcomes, may further help elucidate the relationship between built environment characteristics and health. However, longitudinal neighborhood characteristics and geotagged health outcome data have limited availability, which continue to hinder research on neighborhood effects.

5. Conclusions

The characteristics of the places where we live, learn, work, play, and pray can impact our health. In this study, we harness the underutilized potential of street image data to create a national dataset of built environment characteristics. Our investigation of the impact of built neighborhood characteristics on health suggests indicators of infrastructure development may be connected with lower chronic disease and premature mortality—but also a modest increase in excessive drinking. While this study found more rural environments were

Table 5
Google Street View-derived predictors of census tract health outcomes^a.

	Obesity	Diabetes	Physical distress	Mental distress	Physical inactivity	Binge drinking	Limited access to healthy food	Dental care
	Prevalence difference (95% CI) ^a	Prevalence difference (95% CI) ^a	Prevalence difference (95% CI) ^a	Prevalence difference (95% CI) ^a	Prevalence difference (95% CI) ^a	Prevalence difference (95% CI) ^a	Prevalence difference (95% CI) ^a	Prevalence difference (95% CI) ^a
Google Street View rural area								
3rd tertile (highest)	4.80 (4.48, 5.12) [*]	1.28 (1.11, 1.44) [*]	1.70 (1.53, 1.86) [*]	1.42 (1.29, 1.55) [*]	4.84 (4.50, 5.18) [*]	-1.88 (-2.13, -1.63) [*]	34.48 (32.78, 36.17) [*]	-5.55 (-5.93, -5.17) [*]
2nd tertile	3.39 (3.20, 3.58) [*]	0.76 (0.68, 0.83) [*]	0.81 (0.72, 0.90) [*]	0.41 (0.34, 0.48) [*]	2.77 (2.57, 2.98) [*]	-1.38 (-1.51, -1.25) [*]	23.10 (21.76, 24.45) [*]	-2.78 (-2.99, -2.57) [*]
Census derived								
Population density								
1st tertile (lowest)	2.82 (2.56, 3.07) [*]	0.54 (0.42, 0.67) [*]	0.81 (0.69, 0.94) [*]	0.72 (0.62, 0.81) [*]	2.36 (2.08, 2.65) [*]	-1.04 (-1.21, -0.87) [*]	36.82 (35.76, 37.88) [*]	-3.46 (-3.77, -3.16) [*]
2nd tertile	2.16 (2.04, 2.28) [*]	0.51 (0.46, 0.56) [*]	0.64 (0.58, 0.70) [*]	0.37 (0.32, 0.41) [*]	1.52 (1.39, 1.66) [*]	-1.12 (-1.20, -1.05) [*]	23.20 (22.36, 24.05) [*]	-2.26 (-2.41, -2.12) [*]
Rural census tract	1.72 (1.38, 2.05) [*]	0.22 (0.06, 0.38) [*]	0.55 (0.40, 0.70) [*]	0.66 (0.55, 0.78) [*]	1.65 (1.29, 2.02) [*]	-0.37 (-0.60, -0.15) [*]	22.33 (21.29, 23.38) [*]	-2.55 (-2.93, -2.18) [*]
USDA Rural-urban continuum codes ^b								
Small town & rural (vs. metropolitan tracts)	1.06 (0.92, 1.20) [*]	2.72 (2.64, 2.79) [*]	3.93 (3.84, 4.01) [*]	1.49 (1.43, 1.55) [*]	-1.74 (-1.91, -1.58) [*]	-1.78 (-1.87, -1.68) [*]	32.67 (19.50, 45.84) [*]	-1.44 (-1.64, -1.24) [*]
N	9991	9991	9991	9991	9991	9991	10,529	9991

^a Data source of health outcomes: City Health Dashboard on 500 U.S. Cities. Census tract built environment characteristics categorized into tertiles with the lowest tertile serving as the referent group. Adjusted linear regression models were run for each predictor and outcome separately. Models controlled for census tract-level demographics: population density, rural census tract designation, percent 10–24 years old, percent 65 years and older, percent Hispanic, percent non-Hispanic black, households with relatives (other than spouse and children), households with unmarried partner, owner-occupied housing, economic disadvantage, and household size. A census tract was urban if the geographic centroid of the tract was in an area with > 2500 people; all other tracts are rural. Robust standard errors reported. Separate models were run for each outcome and for each predictor (Google Street View derived rural area, census population density, rural census tract) because the predictors were collinear with each other.

^b Rural-Urban continuum codes: <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx#.U9J07GPDWH0>.

* $p < 0.05$.

characterized by worse health outcomes, the link is not inevitable. Comprehensively promoting health may necessitate tackling multifactorial and structural influences on health—including advocating for roads, community resources, and healthy neighborhood designs—especially in more resource poor areas. More equity in access to health resources may lead to more equity in health outcomes. Neighborhood data can be utilized by public health practitioners, government agencies, city planners, nonprofits, and health care institutions to conduct community risk assessments and inform structural strategies for improving community health.

Acknowledgments

The authors declare no conflicts of interest. This study was supported by the National Institutes of Health's grants 5K01ES025433 and R01 LM012849 (PI: Nguyen, QC); K99MD012615 (PI: Nguyen, T.). This research benefited from the use of credits from the National Institutes of Health (NIH) Cloud Credits Model Pilot, a component of the NIH Big Data to Knowledge (BD2K) program. We thank Weijun Yu for her research assistance.

Appendix A. Supplementary tables and figures

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pmedr.2019.100859>.

References

- Alvarado, S.E., 2016. Neighborhood disadvantage and obesity across childhood and adolescence: evidence from the NLSY children and young adults cohort (1986–2010). *Soc. Sci. Res.* 57, 80–98. <https://doi.org/10.1016/j.ssresearch.2016.01.008>.
- Bader, M.D., Mooney, S.J., Bennett, B., Rundle, A.G., 2017. The promise, practicalities, and perils of virtually auditing neighborhoods using Google Street View. *Ann. Am. Acad. Pol. Soc. Sci.* 669, 18–40.
- Barrientos-Gutierrez, T., et al., 2017. Neighborhood physical environment and changes in body mass index: results from the multi-ethnic study of atherosclerosis. *Am. J. Epidemiol.* 186, 1237–1245. <https://doi.org/10.1093/aje/kwx186>.
- Beifort Christie, A., Nazir, N., Perri Michael, G., 2012. Prevalence of obesity among adults from rural and urban areas of the United States: findings from NHANES (2005–2008). *J. Rural. Health* 28, 392–397. <https://doi.org/10.1111/j.1748-0361.2012.00411.x>.
- Black, J.L., Macinko, J., Dixon, L.B., Fryer, J.G.E., 2010. Neighborhoods and obesity in New York City. *Health Place* 16, 489–499. <https://doi.org/10.1016/j.healthplace.2009.12.007>.
- Block, J.P., Scribner, R.A., DeSalvo, K.B., 2004. Fast food, race/ethnicity, and income. *Am. J. Prev. Med.* 27, 211–217. <https://doi.org/10.1016/j.amepre.2004.06.007>.
- Boothe, V.L., Shendell, D.G., 2008. Potential health effects associated with residential proximity to freeways and primary roads: review of scientific literature, 1999–2006. *J. Environ. Health* 70, 33–41.
- Brookfield, K., Tilley, S., 2016. Using virtual street audits to understand the walkability of older adults' route choices by gender and age. *Int. J. Environ. Res. Public Health* 13, 1061. <https://doi.org/10.3390/ijerph13111061>.
- Chan, L., Hart, L.G., Goodman David, C., 2007. Geographic access to health care for rural medicare beneficiaries. *J. Rural. Health* 22, 140–146. <https://doi.org/10.1111/j.1748-0361.2006.00022.x>.
- Chen, S.E., Florax, R.J., Snyder, S.D., 2013. Obesity and fast food in urban markets: a new approach using geo-referenced micro data. *Health Econ.* 22, 835–856.
- Clarke, C.A., et al., 2010. Racial and social class gradients in life expectancy in contemporary California. *Soc. Sci. Med.* 70, 1373–1380. <https://doi.org/10.1016/j.socscimed.2010.01.003>.
- Clifton, K.J., Livi Smith, A.D., Rodriguez, D., 2007. The development and testing of an audit for the pedestrian environment. *Landsc. Urban Plan.* 80, 95–110. <https://doi.org/10.1016/j.landurbplan.2006.06.008>.
- Day, K., Boarnet, M., Alfonso, M., Forsyth, A., 2006. The Irvine–Minnesota inventory to measure built environments: development. *Am. J. Prev. Med.* 30, 144–152.
- Eames, M., Ben-Shlomo, Y., Marmot, M.G., 1993. Social deprivation and premature mortality: regional comparison across England. *BMJ* 307, 1097–1102.
- Eberhardt, M.S., Pamuk, E.R., 2004. The importance of place of residence: examining health in rural and nonrural areas. *Am. J. Public Health* 94, 1682–1686.
- Eberhardt, M.S., et al., 2001. Urban and rural health chartbook. In: *Health, United States*. vol. 2001.
- Egan, M., Petticrew, M., Ogilvie, D., Hamilton, V., 2003. New roads and human health: a systematic review. *Am. J. Public Health* 93, 1463–1471.
- Evans, G.W., 2003. The built environment and mental health. *J. Urban Health* 80, 536–555.
- Ewing, R., Handy, S., Brownson, R.C., Clemente, O., Winston, E., 2006. Identifying and measuring urban design qualities related to walkability. *J. Phys. Act. Health* 3, S223.
- Farber, D., 2012. CNET.
- Feuillet, T., et al., 2016. Neighbourhood typology based on virtual audit of environmental obesogenic characteristics. *Obes. Rev.* 17, 19–30. <https://doi.org/10.1111/obr.12378>.
- Frank, L.D., Andresen, M.A., Schmid, T.L., 2004. Obesity relationships with community design, physical activity, and time spent in cars. *Am. J. Prev. Med.* 27, 87–96. <https://doi.org/10.1016/j.amepre.2004.04.011>.
- Galea, S., Ahern, J., Rudenstine, S., Wallace, Z., Vlahov, D., 2005. Urban built environment and depression: a multilevel analysis. *J. Epidemiol. Community Health* 59, 822–827.
- Grigsby-Toussaint, D.S., et al., 2010. Neighborhood socioeconomic change and diabetes risk: findings from the Chicago childhood diabetes registry. *Diabetes Care* 33, 1065–1068. <https://doi.org/10.2337/dc09-1894>.
- Hankey, S., Marshall, J.D., Brauer, M., 2011. Health impacts of the built environment: within-urban variability in physical inactivity, air pollution, and ischemic heart disease mortality. *Environ. Health Perspect.* 120, 247–253.
- Hara, K., Le, V., Froehlich, J., 2013. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 631–640 (ACM).
- Hartley, D., 2004. Rural health disparities, population health, and rural culture. *Am. J. Public Health* 94, 1675–1678.
- Heinrich, K.M., et al., 2008. How does the built environment relate to body mass index and obesity prevalence among public housing residents? *Am. J. Health Promot.* 22, 187–194.
- Hembree, C., et al., 2005. The urban built environment and overdose mortality in New York City neighborhoods. *Health Place* 11, 147–156.
- Hyam, R., 2017. Automated image sampling and classification can be used to explore perceived naturalness of urban spaces. *PLoS One* 12, e0169357. <https://doi.org/10.1371/journal.pone.0169357>.
- Kelly, C.M., Wilson, J.S., Baker, E.A., Miller, D.K., Schootman, M., 2012. Using Google street view to audit the built environment: inter-rater reliability results. *Ann. Behav. Med.* 45, S108–S112.
- Khan, L.K., et al., 2009. Recommended community strategies and measurements to prevent obesity in the United States. *Morb. Mortal. Wkly. Rep. Recomm. Rep.* 58, 1–29.
- Kim, J.J., et al., 2004. Traffic-related air pollution near busy roads: the East Bay Children's respiratory health study. *Am. J. Respir. Crit. Care Med.* 170, 520–526.
- Less, E.L., et al., 2015. Matching study areas using Google street view: a new application for an emerging technology. *Eval. Program Plann.* 53, 72–79.
- Li, F., et al., 2008. Built environment, adiposity, and physical activity in adults aged 50–75. *Am. J. Prev. Med.* 35, 38–46.
- Li, X.-j., et al., 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban For. Urban Green.* 14, 675–685. <https://doi.org/10.1016/j.ufug.2015.06.006>.
- Lysy, Z., et al., 2013. The impact of income on the incidence of diabetes: a population-based study. *Diabetes Res. Clin. Pract.* 99, 372–379. <https://doi.org/10.1016/j.diabres.2012.12.005>.
- Meit, M., et al., 2014. The 2014 Update of the Rural-Urban Chartbook. Rural Health Reform Policy Research Center, Bethesda, MD.
- Mooney, S.J., et al., 2014. Validity of an ecometric neighborhood physical disorder measure constructed by virtual street audit. *Am. J. Epidemiol.* 180, 626–635.
- Morris, J.N., Blane, D.B., White, I.R., 1996. Levels of mortality, education, and social conditions in the 107 local education authority areas of England. *J. Epidemiol. Community Health* 50, 15–17.
- Movshovitz-Attias, Y., et al., 2015. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 1693–1702.
- Mujahid, M.S., et al., 2008. Relation between neighborhood environments and obesity in the multi-ethnic study of atherosclerosis. *Am. J. Epidemiol.* 167, 1349–1357. <https://doi.org/10.1093/aje/kwn047>.
- Naik, N., Philipoom, J., Raskar, R., Hidalgo, C., 2014. Streetscore-predicting the perceived safety of one million streetscapes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 779–785.
- Parks, S.E., Housemann, R.A., Brownson, R.C., 2003. Differential correlates of physical activity in urban and rural adults of various socioeconomic backgrounds in the United States. *J. Epidemiol. Community Health* 57, 29.
- Rundle, A.G., Bader, M.D.M., Richards, C.A., Neckerman, K.M., Teitler, J.O., 2011. Using Google street view to audit neighborhood environments. *Am. J. Prev. Med.* 40, 94–100. <https://doi.org/10.1016/j.amepre.2010.09.034>.
- Saelens, B.E., Handy, S.L., 2008. Built environment correlates of walking: a review. *Med. Sci. Sports Exerc.* 40, S550–S566. <https://doi.org/10.1249/MSS.0b013e31817c67a4>.
- Sallis, J.F., et al., 2018. Neighborhood built environment and socioeconomic status in relation to physical activity, sedentary behavior, and weight status of adolescents. *Prev. Med.* 110, 47–54. <https://doi.org/10.1016/j.ypmed.2018.02.009>.
- Silva, V., Grande, A., Rech, C., Peccin, M., 2015. Geoprocessing via Google Maps for assessing obesogenic built environments related to physical activity and chronic noncommunicable diseases: validity and reliability. *J. Healthc. Eng.* 6, 41–54.
- Townsend, P., Phillimore, P., Beattie, A., 1988. Health and Deprivation. Inequality and the North. Routledge.
- TRIP, 2017. Rural Connections: Challenges and Opportunities in America's Heartland.

- Washington, DC. .
- Truong, K.D., Ma, S., 2006. A systematic review of relations between neighborhoods and mental health. *J. Ment. Health Policy Econ.* 9, 137–154.
- Tyroler, H.A., Wing, S., Knowles, M.G., 1993. Increasing inequality in coronary heart disease mortality in relation to educational achievement: profile of places of residence, United States, 1962–87. *Ann. Epidemiol.* 3, S51–S54.
- U.S. Census Bureau, 2015. American Community Survey: 2011–2015.
- Wilcox, S., Castro, C., King, A.C., Housemann, R., Brownson, R.C., 2000. Determinants of leisure time physical activity in rural compared with urban older and ethnically diverse women in the United States. *J. Epidemiol. Community Health* 54, 667.
- Wing, S., Barnett, E., Casper, M., Tyroler, H.A., 1992. Geographic and socioeconomic variation in the onset of decline of coronary heart disease mortality in white women. *Am. J. Public Health* 82, 204–209.
- Yin, L., Wang, Z., 2016. Measuring visual enclosure for street walkability: using machine learning algorithms and Google street view imagery. *Appl. Geogr.* 76, 147–153. <https://doi.org/10.1016/j.apgeog.2016.09.024>.
- Yin, L., Cheng, Q., Wang, Z., Shao, Z., 2015. Big data' for pedestrian volume: exploring the use of Google Street View images for pedestrian counts. *Appl. Geogr.* 63, 337–345. <https://doi.org/10.1016/j.apgeog.2015.07.010>.